

The gathering storm

Hurricane Katrina, which struck New Orleans last year, has thrust the link between climate change and extreme weather events onto the US political agenda.

The hurricane season officially opens today in the United States, although the aftershocks from last August's Hurricane Katrina have yet to die down. The breached levees of New Orleans have been repaired, and the government has stockpiled food and water for a million potential storm refugees, but the political structures that mismanaged Katrina remain disturbingly intact.

The now-notorious Federal Emergency Management Agency (FEMA) remains unconvincingly appended to the dysfunctional Department of Homeland Security. And the voters of New Orleans, many still scattered across the country, managed last month to re-elect Mayor Ray Nagin, the individual most closely associated with the city's woeful failure to prepare itself for the storm.

Bad as Katrina was for the city, it could have been even worse. Max Mayfield, director of the National Hurricane Center, points out that New Orleans itself experienced only the winds of a category-1 hurricane. Yet more than 1,500 Louisianans died, in a flooding disaster that had been predicted for decades. Before Katrina struck, several opportunities for science to properly inform public policy were missed. Afterwards, many asked how a disaster that had been so precisely foreseen could have been allowed to happen.

But can science help now? Can it be integrated into the extraordinary task of rebuilding, planning and growing an entire coast? The answer remains distressingly unclear.

This summer, several assessments of the episode are being released. A group of engineers funded by the National Science Foundation has just issued its final draft report on why the protective levees surrounding New Orleans failed (see page 556). It goes beyond technical assessment and has harsh words for the organizational breakdowns that led to the Katrina disaster.

Also forthcoming is a report from the US Army Corps of Engineers, which built the waterworks of southern Louisiana. Another is expected from the state's Team Louisiana, led by Ivor van Heerden of Louisiana State University, a hurricane specialist who pre-Katrina served as the loudest Cassandra of all. It is to be hoped that these evaluations will be digested and acted on by those responsible for the reconstruction of New Orleans.

Another contentious issue brought to the fore by Katrina is the question of whether climate change is making hurricanes worse. In the past year, an emerging consensus has suggested that rising sea surface temperatures may well be causing hurricanes to become more intense over time (see page 564).

It can be debated how significant this effect is: will a few degrees of warming cause enough increase in intensity to ruin people's lives? After all, it is the trajectory of a hurricane that most crucially determines how many people suffer. Yet computer models suggest that the rise in intensity will be sufficient to cause concern.

Hurricane experts must take care not to overplay this link. The public is prone to misunderstanding subtle climate connections. Many people are more likely to recall dramatic scare stories about the possible connection than they are scientists' careful caveats that no particular storm can be attributed to global warming.

More worryingly, the science of hurricanes and global warming seems to be falling into the same trap that has ensnared climate-change research for two decades. Researchers are lining up into distressingly familiar camps, with some arguing for the link between tropical storms and climate change, and some against it. They duel at press conferences and snipe at each other on the Internet and in the literature, each side trying to dissect the other's data.

Scientists who have little or no background in hurricanes are suddenly publishing papers in tropical meteorology, sometimes apparently pursuing an agenda. But the community of real hurricane experts is extraordinarily small and doesn't need to be artificially enlarged by people looking to prove a particular point.

As US federal and state agencies prepare to respond to this summer's hurricanes, science has an opportunity to inform and improve the lives of millions of people. Researchers must do their best to ensure that the data, and not the politics, determine the shape of these preparations. ■

"As US agencies prepare to respond to this summer's hurricanes, science has an opportunity to improve the lives of millions of people."

Finding fraud in China

As Chinese research expands, who is looking out for faked results?

The investigation of research misconduct is always fraught with difficulty, even if the necessary protocols and experienced expert committees are fully in place. In China, they are not. If the nation is to get to grips with the problem of misconduct as it

becomes a substantial scientific power, that situation has to change.

Chinese research agencies do have structures for investigating misconduct allegations, but in the absence of open discussion and independent press scrutiny, few researchers have much faith in them. The rapid and open exchange of information over the Internet has some potential to fill the void, but it also carries risks (see *Nature* **441**, 392–393; 2006). It could readily break down into a dangerous game of unregulated accusation and counter-accusation, shedding no light on actual misconduct.

The power of the Internet in identifying scientific fraud was



amply demonstrated last year in the case of Woo Suk Hwang, the discredited South Korean cloning researcher. Online portals discussed suspicious images and data in Hwang's papers, ultimately leading Seoul National University to pursue an investigation that exposed Hwang's fabrications. And Internet postings of allegations that Jin Chen faked digital-processing chips contributed to his dismissal from Shanghai Jiaotong University last month.

The Internet can play a particularly important role in countries such as China and South Korea that do not have adequate systems for investigating misconduct allegations. That isn't to say that countries with systems in place are totally on top of the problem, but at least they have developed some of the institutions and protocols needed to handle it.

Organizations charged with assessing allegations of scientific misconduct do exist in China, and on paper the system appears functional — but there is no evidence that it really works. China lacks an independent press to report on such matters. The very size of the country and subsequent disparate implementation of policies set in Beijing make matters worse.

In addition, the cultural importance of 'saving face' in Chinese society makes the full-frontal public attacks that tend to characterize Western misconduct allegations almost unthinkable. There are no effective provisions to protect whistleblowers, so it is hard to believe that anyone who observes misconduct would summon the courage to report it to the authorities.

It is in this climate that New Threads, a Chinese-language Internet site run by a single researcher based in San Diego, has come to play a significant role in the monitoring of scientific conduct. This

arrangement is deeply problematic, however.

In China's recent history, 'bottom up' accusations have often been abused by the authorities to persecute perceived enemies of the state. This was especially true during the Cultural Revolution, when simply pasting a poster on the wall calling someone a 'bourgeois' could destroy their livelihood. The threat of innocent people being branded as 'pseudoscientists', either by a jealous rival or by the state, further clouds the misconduct picture in China.

The only real solution to this problem is a great deal more complex than hooking up to an Internet connection. It requires the establishment of independent offices in Chinese research agencies, rather like the inspector general's office at the US National Science Foundation, or the Office of Research Integrity at the US health department. The system can only operate effectively if it offers protection to whistleblowers. It also requires a new generation of scientists to be educated in what constitutes proper scientific conduct. And it needs to ensure that investigations give anyone accused the opportunity to demonstrate their innocence.

China is struggling to come to terms with these kinds of requirements in society at large, as well as within the scientific community. For a multiplicity of reasons — of which the desire for scientific progress is just one — addressing them ought to be the government's greatest priority. ■

"There are no effective provisions to protect whistleblowers, so it is hard to believe that anyone observing misconduct would summon the courage to report it."

Last rights

Researchers have a duty to use the most humane means available of killing laboratory animals.

Scientists who work with animals generally agree that if their subjects have to be killed, the death should be accompanied by as little fear or pain as possible. In practice, many laboratory mice and rats are currently killed with carbon dioxide gas. The method is cheap, easy, is what everyone else does, and seems painless enough. The animals are placed in a box, the gas is turned on, and the researcher can walk away.

Some veterinarians and ethicists are not convinced the method is humane, however. Although the sparse amount of investigation done on the topic is inconclusive, they consider that exposure to slowly rising levels of carbon dioxide engenders in rodents the same feeling it does in humans: blind panic. Alternatives exist and, they say, should be used (see page 570).

The main alternative to carbon dioxide, in places where large numbers of animals are involved, is to gas rodents with high doses of anaesthetics used on humans, such as halothane or isoflurane, as this would probably cause the animals little distress. Individual rodents can be dispatched with a manual manoeuvre known, somewhat euphemistically, as cervical dislocation, which breaks the animal's neck. Researchers have been doing this to rats and mice for decades

and, done right, it is instantaneous and painless. Skill can be gained by practising on anaesthetized animals. Anaesthesia and cervical dislocation can also be used in combination.

In the United States, the institutional animal use and care committees that oversee animal research are likely to demand a "scientific justification" for breaking rodents' necks rather than gassing them. This is because most of them adhere to guidelines written by the American Veterinary Medical Association (AVMA), which requires such justification, noting that cervical dislocation may be "aesthetically displeasing to personnel" who have to carry it out. This piece of red tape often makes it simpler for US researchers to stick with carbon dioxide.

Some might argue that scientists who kill mice for their research should be willing to do so with their bare hands, if that is the most humane thing to do. But there is no guarantee that it always will be, and sometimes research regulations, or the particulars of the experiment, may call for the use of one method over another. Cost is another consideration, as anaesthetic gas is more expensive, and often less readily to hand, than carbon dioxide. But everyone who makes these final decisions has a duty to carefully examine the options that are available — and to take into account what's best for the rodent, as well as for the researcher. ■

"The main alternative to carbon dioxide, where large numbers of animals are involved, is to use high doses of anaesthetics used on humans."

RESEARCH HIGHLIGHTS

A change of weather

Nature Phys. doi:10.1038/nphys314 (2006)

The physics that describes the onset of tropical downpours shows intriguing similarities to theories of magnetism, physicists in the United States report.

Some materials become magnetic only when they are cooled below a critical temperature: their magnetism arises as the spins of the material's electrons line up. This change is described as a continuous phase transition.

Ole Peters of the Santa Fe Institute in New Mexico and David Neelin of the University of California, Los Angeles, used satellite data to analyse the relationship between water vapour and rainfall. The weather system undergoes a phase transition mathematically equivalent to that in magnetized materials when rising vapour levels trigger rainfall — a surprising result given the difference in scale between the two systems.



M. S. YAMASHITA/CORBIS

CANCER BIOLOGY**The power of p53**

Science doi:10.1126/science.1126863 (2006)

Researchers in the United States have linked the tumour-suppressor gene *p53* with an oddity of cancer cells that was first noted in the 1930s, known as the Warburg effect.

Normal cells mostly make energy by aerobic respiration, using oxygen to break down sugars. The Warburg effect describes cancer cells' preference for anaerobic breakdown of sugars, a process known as glycolysis.

Paul Hwang at the National Institutes of Health in Bethesda, Maryland, and his colleagues show that *p53* plays a part in shifting the balance by modulating a

mitochondrial protein that regulates oxygen use in the cell. Mouse and human cell lines deficient in either *p53* or mitochondrial protein activity were found to favour anaerobic over aerobic metabolism.

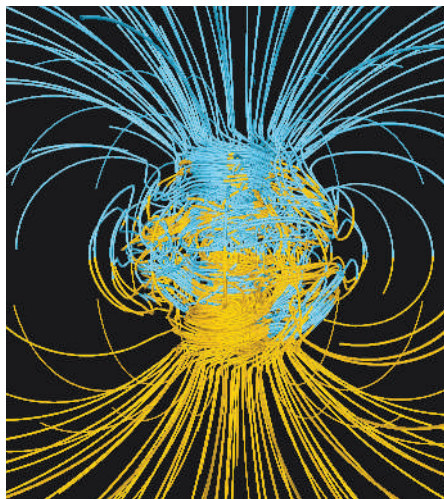
p53's new-found link to respiration may also offer clues to the gene's role in ageing.

EARTH SCIENCE**Magnetic flip**

Earth Planet. Sci. Lett. doi:10.1016/j.epsl.2006.03.038 (2006)

Earth's magnetic field (pictured left) is prone to flip: north and south have swapped, on average, every half-million years for the past 160 million years. But the fact that it's been longer since the last reversal doesn't mean that one is overdue, argue Catherine Constable of the University of California, San Diego, and Monika Korte of Germany's National Research Centre for Geosciences in Potsdam.

Some researchers have speculated that the weakening of Earth's magnetic field since measurements began in the mid-nineteenth century points to an imminent reversal. Constable and Korte counter that both archaeomagnetic data and palaeomagnetic records retrieved from sediment cores show the recent rate of change to be nothing unusual. The team also notes that the current field strength is still higher than the geological average and that stronger fields have, in the past, been associated with longer reversal periods.

**QUANTUM COSMOLOGY****Knots in space**

Phys. Rev. Lett. **96**, 180604 (2006)

The rapid expansion and cooling of the very early Universe may have created knot-like 'topological defects' in the fabric of space-time — a process that has now been mimicked in the lab using a pair of niobium rings.

When rapidly cooled, the niobium rings plunge into a superconducting state that can contain quantum-mechanical defects, analogous to cosmological defects, known as vortices. Roberto Monaco of the University of Salerno in Italy and his co-workers have examined how the probability of vortex formation depends on the cooling rate. They find that the observed relationship matches the predictions made in the 1980s for cosmic topological defects.

PHYSIOLOGY**Hormonal pain**

J. Neurosci. **26**, 5777–5785 (2006)

A brain-imaging study provides new evidence of oestrogen's role in pain perception.

Jon-Kar Zubieta of the University of Michigan in Ann Arbor and his colleagues have previously linked sex differences in pain perception to the regulation of μ -opioid receptors. This receptor system, when activated by sustained pain, mediates the body's natural pain-killing mechanisms.

Now Zubieta and his team have used positron emission tomography and a label of

μ -opioid receptors to examine eight women's responses to pain in certain regions of the brain before and after treatment with oestradiol, a form of oestrogen. They conclude that oestradiol increases the number of μ -opioid receptors and enhances their activation. Subjects that had been treated with oestradiol also reported less pain sensation.

CHEMISTRY

Carbon united

Angew. Chem. Int. Edn doi:10.1002/anie.200600951 (2006)

Chemists have used electricity to make carbon-carbon bonds, a crucial step in building organic molecules, in a microreactor.

In doing so, Stephen Haswell of the University of Hull, UK, and his co-workers combined two beneficial strategies: electrochemistry unites carbon atoms under much milder conditions than are used in more typical metal-catalysed reactions; and microreactors offer a simple route to continuous reactions, carried out as molecules flow through tiny channels.

The researchers reacted various benzyl bromides with a series of alkenes, achieving high yields of products useful in making compounds such as antibiotics.

SYSTEMS NEUROSCIENCE

A touching story

Neuron 50, 617-629 (2006)

Biologists in Switzerland have watched in real time how the behaviour of mice affects how their brains process sensory signals coming from their whiskers.

Carl Petersen and his team at the Swiss Federal Institute of Technology, Lausanne, used voltage-sensitive dyes and a lightweight fibre-optic bundle to watch patches of neurons in the part of the brain that processes whisker signals. Usually mice are anaesthetized before their neuronal activity is measured, but this technique works on mice that are awake and free to move.

The researchers showed that the neurons in still mice responded much more strongly to a whisker touch than the neurons in mice that were swishing their whiskers back and forth.

MATERIALS SCIENCE

Tight squeeze

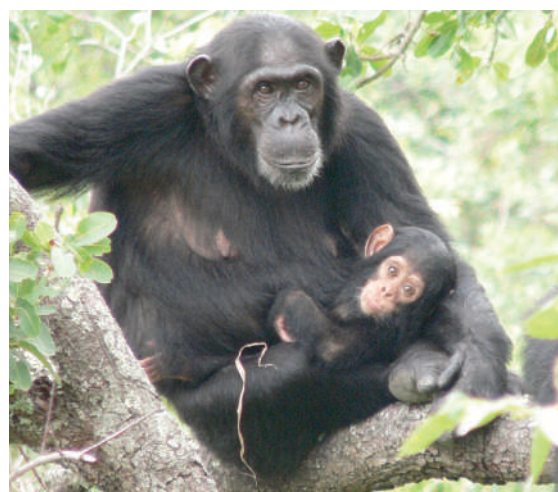
Science 312, 1199-1202 (2006)

Appl. Phys. Lett. 88, 193121 (2006)

Metals stored inside carbon nanotubes can be squeezed out like toothpaste from a tube, according to Florian Banhart of the University of Mainz, Germany, and his co-workers.

Banhart had previously shown that 'nano-onions' — multilayered concentric spheres of graphitic carbon — shrink when electron irradiation kicks out carbon atoms. Now his team reports that irradiating nanotubes filled with metal causes a toothpaste-like extrusion of their contents. The researchers estimate that the pressure created as the nanotubes contract may reach more than a tenth of that at the Earth's core.

In separate experiments, Banhart and his colleagues show how a nano-onion can be used as a high-pressure reaction cell: the internal pressure of the irradiated structure helps carbon atoms to combine with encapsulated iron, forming iron carbide.



M. WILSON

VIROLOGY

Chimps harbour HIV in wild

Science doi:10.1126/science.1126531 (2006)

Wild chimpanzees (pictured above) have long been suspected as the source of the human immunodeficiency virus, HIV. Now that theory has been confirmed: evidence from southern Cameroon shows that wild members of the chimpanzee subspecies *Pan troglodytes troglodytes* harbour the strain of simian immunodeficiency virus (SIV) that is most closely related to pandemic HIV.

Researchers led by Beatrice Hahn of the University of Alabama at Birmingham report the presence of antibodies against the SIV strain in droppings left by wild chimps. This strain had previously only been detected in captive animals.

The virus, which does not seem to cause disease in the chimps, was probably transmitted to humans by exposure to blood during bushmeat hunting or butchery. The location of the discovery also supports the idea that the HIV pandemic had its roots in nearby Kinshasa, in the Democratic Republic of Congo.

JOURNAL CLUB

Daniel Lieberman

Harvard University, Cambridge, Massachusetts, USA

A biological anthropologist recommends reading about optimal running, but advises against trying it.

As someone interested in human locomotion, I frequently watch people run and walk, both on a treadmill in my lab, and on the streets of Cambridge, where I live.

It is well known that humans are good at minimizing cost: we typically walk near our optimal speed in terms of oxygen cost, and switch gaits at approximately the speed for which running costs less than walking.

I often wonder how people gauge what speed to choose, and when to switch gaits. Do we base our behaviour on speed itself, the determinants of speed, such as stride length and stride frequency, or some other variables?

John Bertram of the University of Calgary in Canada and his colleagues (Gutmann *et al.* *J. Exp. Biol.* 209, 622-632; 2006) recently tackled this problem — at least the running part of it — in a fun, creative experiment. They asked athletes to run at particular speeds (by using a treadmill), stride lengths (by making them land on spaced markers in a field) and stride frequencies (by making them time their footstrikes to a metronome). They then used published data to estimate the oxygen costs of the different gaits.

Interestingly, runners compensated for the constrained (and sometimes odd) gaits by adjusting the unconstrained parameters in ways that apparently minimized their oxygen intake. So even when forced to take mincing little steps, or to galumph slowly, the athletes somehow optimally adjusted their speed, stride length and/or stride rate.

How bodies figure out this trick remains a mystery. All I can say is that I've tried constraining my gait in similar ways when I run in the morning, but instead of feeling optimal, I mostly feel foolish.

NEWS

Pandemic 'dry run' is cause for concern

AP PHOTO/BINSAR BAKKARA

A cluster of avian flu cases in Indonesia last month is being seen by many experts as a dry run for the handling of an emerging pandemic virus. But although the World Health Organization (WHO) says that all went well, some critics allege that the response to the virus — thought to have been moving between humans — shows how ill-prepared the international community and affected nations still are.

"Any chance of containment was absolutely hopeless," says Andrew Jeremijenko, who until March was head of influenza surveillance at the US Naval Medical Research Unit 2 in Jakarta. "If this was a test to see whether Indonesia could contain a virus, then they just failed miserably."

The difficulties encountered also raise questions as to the practicality of a plan to try to stop an emerging pandemic in its tracks by rapid intervention. Modelling studies predict that if a pandemic virus emerges, the WHO would have at most three weeks to help the affected country to quarantine all carriers and treat those infected with antivirals (N. M. Ferguson *et al. Nature* **436**, 614–615; 2006).

The first case in the cluster fell ill on 24 April and died on 4 May. Samples were not taken, however, and alarm bells only rang when her relatives started going to hospitals in the days that followed. In total, eight members of an extended family in the village of Kubu Sembelang in north Sumatra became infected with H5N1. Six more of them have since died.

Jeremijenko says the response was slow and disorganized. The first WHO official and a team of local officials didn't reach the village until 12 May. Other international experts did not reach the village until the following week, at least in one case because of difficulties getting an invitation from Indonesia's ministry of health, according to Jeremijenko. Villagers also refused to cooperate with the team initially. Several of the H5N1 patients fled hospitals, returning coughing to the community.

Spreading the news

The WHO made the outbreak public on 18 May. Health officials — and stock markets — worldwide trembled five days later when the WHO budged from its previous standard line that "the most plausible source" of the cluster was infected poultry, and acknowledged for the first time since the emergence of H5N1 that an extended chain of human transmission was the most likely explanation.

Steven Bjorge, a WHO official in Jakarta, disputes the allegation of unnecessary delays



Johannes Ginting is thought to have caught bird flu from a relative. Seven members of his family have died.

and bungling, arguing that the WHO and the Indonesian government reacted promptly. "The team was in the field early, and the Indonesians are doing a good job," he says. The abscondments from hospital were "an unusual experience," he adds.

Concerns over the cluster itself have eased as no new cases have since been reported nearby, and the WHO says the virus's sequence shows "no evidence of significant mutations". The sequences have not been made public yet. The all-clear will not be given for another two weeks or so, however, and the pharmaceutical company Roche has been put on standby to send antiviral drugs to the region.

Teams on the ground are trying to monitor fresh cases. But thousands of Indonesians die every day from tuberculosis, dengue and other infectious diseases, and almost all go untested

for H5N1. On 29 May, the WHO announced six more cases in other areas of Indonesia, two of which were also a family cluster.

"There have been a number of family clusters where only one person was tested," says Jeremijenko, adding that there is "only limited

testing, in large cities such as Surabaya, Medan, Bandung and Jakarta. We know we are missing cases, especially in rural areas."

What caused the suspected human-to-human transmission at Kubu Sembelang is still a mystery. *Nature* has learned that the cases differed from past Indonesia cases, in that they had much higher viral loads in the throat and nose. Human-to-human transmission is more likely through droplets coughed from the nose and throat than from infections further down the respiratory tract.

Mutations in cases in Turkey earlier this

"If this was a test to see whether Indonesia could contain a virus, they failed miserably."



JAVA EARTHQUAKE
Find *Nature's* coverage of this disaster on our website
www.nature.com/news

AP PHOTO/PURWOWIYOTO

Murders halt rainforest research

Following the murder of two guards by illegal goldminers, scientists have been evacuated from a research station in the Nouragues nature reserve in French Guiana. The assassins were caught last week in Régina, a village lying 40 km from the station.

The bodies of Domingo Ribamar da Silva and Andoe Saaki ('Capi') were found on 18 May at the Arataï ecotourism centre where they worked, a few kilometres from the Saut Pararé and Inselberg research sites — all run by France's basic research agency, the CNRS. They had been shot, and the centre ransacked.

Alain Pavé, the head of CNRS in French Guiana, ordered the temporary closure of the research station the day after the discovery of the bodies. Fourteen staff and their equipment were evacuated by helicopter to Cayenne, 100 km away. The gendarmerie is protecting the few remaining staff, and the sites' facilities. The research station will not be reopened until security can be guaranteed, says Pavé.

"Apparently this murder is deliberate, probably to chase the scientists and managers of the Nouragues reserve from the area," says Pierre Charles-Dominique, the tropical ecologist who heads the research station.

The miners, or 'garimpeiros', most of whom come from neighbouring Brazil, have been a constant security concern for scientists, he adds. His station was ransacked by miners in 2004. The thieves made off with equipment worth €75,000 (US\$100,000), and delayed a major research project (see *Nature* **430**, 127; 2004).

Illegal goldmining is a growing problem in the reserve, says Patrick Jansen, an ecologist at the University of Groningen in the Netherlands. The miners cause substantial ecological damage: they suction riverbeds for gold nuggets, fell trees, and pollute waterways with mud and mercury.

Before the recent murders, scientists at the reserve were preparing to celebrate completion of the project delayed by the 2004 ransacking. The Canopy Operation Permanent Access System (COPAS) is designed to give scientists unprecedented access to the canopy of a tropical rainforest. It consists of a huge helium balloon and basket, which allows two scientists at a time to move vertically and horizontally within the canopy along a system of cables. Domingo and Capi had lunched with the COPAS research team at Saut Pararé the Saturday before they were killed, researchers recall, as they put the final touches to the treetop system.

All scientific work has now been put on hold. After the 2004 ransacking, the gendarmerie stepped up actions against the miners, introducing helicopter missions to find and destroy the mining sites, and arrest the miners, for example. But Charles-Dominique says that

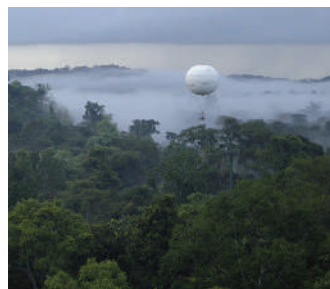
is not enough. "We cannot continue our research without a permanent military guard at the entrance of the reserve, and strong repressive action throughout the park," he says. And that's no easy task in the vast forest.

Catherine Bréchignac, president of the CNRS, visited the centre on 25 May, as part of a scheduled visit to inaugurate COPAS. In response to the fears of some researchers that the station might be closed, she says "certainly not!". After discussions with local government, she promised that the centre will be protected by the gendarmes and the military until the CNRS finds a private firm of armed security guards to assist it.

"I made the decision last night: two murders is inadmissible, we can't let visiting scientists work under such conditions," she says. "I don't know much it will cost, but the science here is very important. The CNRS is committed to it and we will pay whatever it takes." As *Nature* went to press, scientists were due to return to the station on 30 May, and COPAS will now be officially launched next week.

Still, the incident has left researchers badly shaken. Ecologist Pierre-Michel Forget worked at the station from 1984 until 2003, when he decided that the situation was too dangerous and returned to his native France. "Not everyone agreed with me in the scientific community," he says. "But today, unfortunately, I see that I was right to do so. We often stayed at the Arataï river camp when reaching the forest and the station by boat. This could have been me."

Declan Butler



A balloon will give scientists access to the forest canopy once work resumes.

R. LEGUEN



year showed a substitution of glutamic acid with lysine at position 627 in the PB2 component of the polymerase gene. The mutation is thought to allow the virus to survive in the cooler nasal regions. This mutation has not been publicly reported in Indonesia previously, but *Nature* has learned that it occurred in at least one case in August 2005.

Another explanation is that the first case in last month's cluster had a particular genetic susceptibility to the virus, making her a 'super-spreader'. But it is too soon and the data are too sparse to know for sure, says Bjorge.

Malik Peiris, a virologist at the University of Hong Kong who sequenced the virus, declined to comment on any mutations, saying that making sequences public is not his call. "Our job as a WHO reference lab is to report back to the originating country and the WHO," he says. The WHO also declined to give any details. "We will leave that to the government of Indonesia, the owner of the data," says Bjorge.

Declan Butler

Q&A

ON THE RECORD

“Carbon dioxide. They call it pollution. We call it life.”

Television adverts from the Competitive Enterprise Institute — a group that receives funds from the oil industry — imply that rising levels of CO₂ are nothing to be alarmed about.

“The ads are a deliberate attempt to confuse and mislead the public.”

Curt Davis of the University of Missouri, Columbia, criticizes the institute for misusing his findings on the East Antarctic ice sheet.

Sources: CEI, St. Louis Post-Dispatch

SCORECARD



Rampaging elephants
Sri Lanka announces plans to tame, rather than kill, unruly wild elephants. It will then use the beasts as tourist attractions.



Race of the clones
The first cloned equines — two mules called Idaho Gem and Idaho Star — are ready to run against each other, and others, in competitive mule races.



Rabbit romance
Urologists develop prosthetic penises for male bunnies. The implants could have human applications, too.



Tracking whales
University of Washington researchers are training failed drug-detecting dogs to sniff out the floating faeces of endangered killer whales.

OVERHYPED

A golf shot round the world

NASA mission planners have delayed a spacewalk to drive a gold-plated golf ball into low Earth orbit. The publicity stunt, sponsored by Canadian golfing firm Element 21, would have made International Space Station Commander Pavel Vinogradov a sporting celebrity. But engineers are worried that Vinogradov's hook could lodge the ball in a solar panel or other piece of vital equipment. NASA wouldn't comment on Vinogradov's handicap, saying only that the swing would be put off for a more thorough safety analysis.

Drilling for truth in New Orleans: a geologist's story

Since Hurricane Katrina hit, David Rogers has travelled to New Orleans 11 times. A geologist from the University of Missouri, Rolla, Rogers has been investigating the failure of the levee system — this collapsed in several places during the storm, flooding the city. On 22 May, Rogers and others in a team funded by the National Science Foundation released a final draft report on the disaster. The report goes beyond explaining the geology and blames penny-pinching for much of the chaos.

What was your role in the team?

I supervised characterization of the failure sites — drilling holes and recreating the foundation conditions. I also had responsibility for historical research on the background of the structures: how the New Orleans flood-protection system has evolved since 1718. Both overwhelming tasks, believe me.

What was it like working in New Orleans?

It was like pictures you see of Hiroshima and Nagasaki after the Second World War. Just complete devastation for mile after mile. No people, no bathrooms, no water. Choking dust; very fine dust on everything.

The drilling was very difficult. A drill rig took six hours to get out of Baton Rouge, through all the traffic. Just horrible.

There were lots of gawkers driving by. And constant media coverage. Sometimes I'd be talking to the grad students and I'd find my words on the front of New Orleans' *Times-Picayune* the next morning. I had so many people asking me to opine on things long

before the analysis had even gotten going. They wanted to know, why? Who do I blame?

What went wrong with the levees?

We don't think that everything was due to overtopping. We feel a lot more of it was seepage related. When the storm hit, water was forced under the structures, eroding their bases and knocking them down.

What was really unusual about drilling down there is how permeable the swamp deposits are. I've drilled in the Dead Sea, I've been all over the world and I have never seen anything like the swamp mucks underneath New Orleans. I would be pushing a tube sampler down one hole and 30 feet away water comes squirting up my old bore-holes.

The flow of water through this permeable ground should be cut off by large metal sheets that extend deep into the ground underneath the levees. But a lot of the sheet piles did not extend far enough. Had those sheet piles been just 5 feet deeper, a lot of the failures wouldn't have happened — that's how close they were.

The thing is, you can't just use one average depth for the whole structure, you have to keep evaluating the foundation every foot. Bad assumptions were made in a few places, where the geology dipped down, and they didn't get the cut-off deep enough.

When we looked at the records we saw an increasing drive towards economy in the past 20 years. It would appear that they tried to do more for less, and they saved money, but they sacrificed a lot on safety.

I found the guy who drew the cross-section that was wrong on 17th Street, I actually tracked him down. He feels horrible. He was



Mammoth task:
David Rogers says New Orleans is not ready for the hurricane season, which starts this month.



New Orleans was devastated after Katrina struck on 29 August 2005. But why was the flooding so bad?

told to draw the line where he did. He says he was told to average it. People were generally hired, fired and promoted based on one thing: whether they brought their projects in on time and on budget.

So who is to blame?

There is plenty of blame at all levels, right up to the executive branch. It is very disappointing that the decision-makers have a difficult time appreciating probabilistic risk assessment. When we tell them, as scientists, that something has a 25–40% chance of happening, that seems like a low chance to a layperson. To us, that seems a huge risk.

In my opinion, key decisions seem to have been made by people whose principal concerns were economy and timing, not safety. The local levy boards and the sewage and water boards down there are constantly at cross-purposes. Instead of working out compromises they engage in costly litigation, and these boards tend to be staffed by political appointees, not people with appropriate technical expertise. On the flight home I'd just think, is there anybody with a brain down there overseeing the big picture? It is scary.

We need real experts to advise the president and Congress on long-term stuff and crisis response. National Guard generals aren't sufficient to respond to a tsunami, earthquake or flood. They don't have experience working with these things.

The hurricane season starts on 1 June. Are we ready?

No, the band-aids won't be completed until 30 June. Not for lack of trying — it was just too much damage. But the whole system needs to be re-evaluated, because it is just going to break someplace else. They've got the break sites covered like a glove, but the drainage capacity of the city is diminished.

Here's what your readers need to know. You don't know what a 100-year storm is unless you have 1,000 years of records. When I was drilling that swamp out there, looking at the sediment disposition was just like going through tree rings. It was one hurricane after another, after another, after another. I couldn't even count them all.

How do you feel, now the work is done?

I'm really discouraged right now, because we did not get much response. As far as I can see, none of the elected officials seemed to care. We only had one show up to our press conference. We had the world's finest minds. We have a well-thought-out plan of attack for how to fix the problems that created this, but we don't have an audience for it.

My wife would tell you that working on this has changed me. I am fighting against becoming overly pessimistic. I am really disappointed by my country. ■

Interview by Emma Marris

The report can be found at

♦ www.ce.berkeley.edu/~new_orleans



**STOMACH BUG MAKES
FOOD RELEASE CALORIES**
Plus more from the
American Society of
Microbiology meeting.
www.nature.com/news

Old tools shed light on hobbit origins

They may have been tiny, but the hobbits of the Indonesian island of Flores are still the focus of the biggest controversy in anthropology. The latest twist in the tale suggests that these one-metre-tall hominids, with a brain the size of a grapefruit, were the final members of a tool-making tradition stretching back more than 800,000 years. But amid fresh doubts over the species' evolutionary history, the idea that the curious creatures were deformed modern humans refuses to go away.

News of the discovery of the 18,000-year-old remains of tiny people amazed the world when it was revealed in October 2004 (P. Brown *et al. Nature* **431**, 1055–1061; 2004; M. J. Morwood *et al. Nature* **431**, 1087–1091; 2004). Researchers named them *Homo floresiensis* after their home, and concluded that they were ancestors of the much larger *Homo erectus* — the only hominid species known to have been in Asia at the right time. Since then, at least eight more individuals have been found (although the original specimen, called LB1, remains the only skull), dating from around 90,000 to just 12,000 years ago (M. J. Morwood *et al. Nature* **437**, 1012–1017; 2005).

Some experts, such as Robert Martin of the Field Museum in Chicago, have never accepted that the fossils represent a new species of hominid, arguing that LB1's brain was just too



Tools from Liang Bua cave, home of *H. floresiensis*, are similar to older stone blades made by *H. erectus*.

small. He believes the hobbits were small modern humans and that LB1 had a deformed, miniature brain — a condition called microcephaly. In this hypothesis, tools found with the hobbits' remains must have been made by normal *Homo sapiens*. A recent exchange (see *Science* **312**, 999; 2006) argued about whether casts of the inside of the hobbit skull resemble that of a microcephalic human — a debate that seems unlikely to be resolved soon (see 'Is the Flores hobbit a deformed *Homo sapiens*?').

In *Nature* this week, a separate line of evidence points to *H. floresiensis* as a tool-maker. More than 500 stone blades found on Flores and dated to more than 700,000 years ago seem to have been made in the same way — by striking stones to chip off large flakes — as the more recent blades found with the hobbits.

The earlier tools were almost certainly made by *H. erectus*, say Adam Brumm of the Australian National University in Canberra and his colleagues (see page 624). They conclude that *H. floresiensis* may have inherited the technique from *H. erectus*, although they admit more evidence is needed, and are heading to Flores this month to try to acquire it.

The *H. erectus* case is also supported by results from Susan Larson of Stony Brook University in New York. In April, she told the meeting of the Paleoanthropology Society in Puerto Rico that *H. floresiensis* has a shoulder joint more similar to *H. erectus* than *H. sapiens*.

However, even the researchers who unveiled the hobbit bones now think that Indonesia's *H. erectus* may have been too big to evolve into something so diminutive. Michael Morwood of the University of New England in Armidale, Australia, says he believes the hobbits may instead descend from a smaller, as-yet-undiscovered hominid, resembling 1.8 million-year-old specimens found at Dmanisi in Georgia.

More fossils and artefacts are needed to settle the dispute. But views about *H. floresiensis* are now so entrenched that it will be difficult for some to retract their arguments, says Chris Stringer, an anthropologist at the Natural History Museum, London: "Whoever digs in Flores now has to do so with an open mind." ■
Michael Hopkin

Is the Flores hobbit a deformed *Homo sapiens*?

YES

Robert Martin, Field Museum, Chicago

The brain is too small to fit anything I

know about. I calculated that if you start with a 60-kilogram *Homo erectus*, you have to go down to 2 kilograms to get a brain that small — the size of a domestic cat. That started me thinking that LB1 must have been microcephalic.

The average brain size for a human microcephalic brain is 400 cm³, the same as LB1. The brain of a 32-year-old microcephalic woman from Lesotho (skull pictured above) shows that, although roughly 75% of microcephalics die before adulthood, some get by surprisingly well.

Although the other Flores individuals seem to have been small, they may have been a community with a microcephalic living among them, in which people with normal brains would have made the tools.



NO

Dean Falk, Florida State University, Tallahassee

Microcephalics can and do have brains this size. But the LB1 brain doesn't have a microcephalic shape. We now also have parts of nine individuals, which shows LB1 is not a single pathological specimen.

The Flores skull also has features that you just don't see in the brain of *Homo sapiens*, particularly in the temporal lobes and the frontal lobe — an area involved in complex thinking. You need to look not just at size, but also organization, the internal wiring. This thing is small, but it's specialized.

Martin's team don't provide enough detail to show the identifying features and landmarks of the brain. And as for the microcephalic human from Lesotho (pictured on left), compare that with LB1 (above). They do not look alike.



M.H.

SPECIAL REPORT

How dangerous is chemistry?

The death of a French professor in a laboratory explosion in March was a shocking reminder that research can be a risky business.

Mark Peplow and **Emma Marris** investigate whether chemistry deserves its reckless reputation.

Something that felt like an earthquake hit the French town of Mulhouse on 24 March. The explosion at the National Institution of Higher Learning in Chemistry (ENSCMu) killed Dominique Burget, a 41-year-old photochemist. It also sent ripples of concern around the world.

Although official investigations are expected to last until the end of the year, it appears that residues of the flammable gas ethene in a pressure vessel were responsible. Burget was working in the lab above the explosion and had nothing to do with the experiment, which also severely injured a 19-year-old student in the room next door. "She is now out of danger and comes back to the school next week,"

Serge Neunlist, director of ENSCMu, said on 23 May. The explosion caused roughly €10 million (US\$130 million) of damage and destroyed about 4,000 m² of the building, which will take at least three years to rebuild.

Chemistry's reputation for big bangs might suggest that Mulhouse was no freak accident. Gather any group of chemists together and before long they are likely to be exchanging stories about heart-stopping near-misses or discussing someone involved in a serious accident.

Is chemistry really so dangerous? Those responsible for the safety of research labs say such stories may perpetuate out-of-date myths. "A lot of it is reminiscence to 'the good old days' of chemistry," says Alan Kendall, safety officer at the University of Oxford, UK.

"There's a public perception that is years behind the reality," agrees Richard Firn, a biologist who chairs the laboratory safety committee at the University of York, UK. "Things have

changed a lot in the past 10 to 15 years."

Swathes of occupational-health legislation in the 1970s, which established, for example, the US government's Occupational Safety and Health Administration (OSHA) and the UK Health and Safety Executive (HSE), have spurred the change in culture.

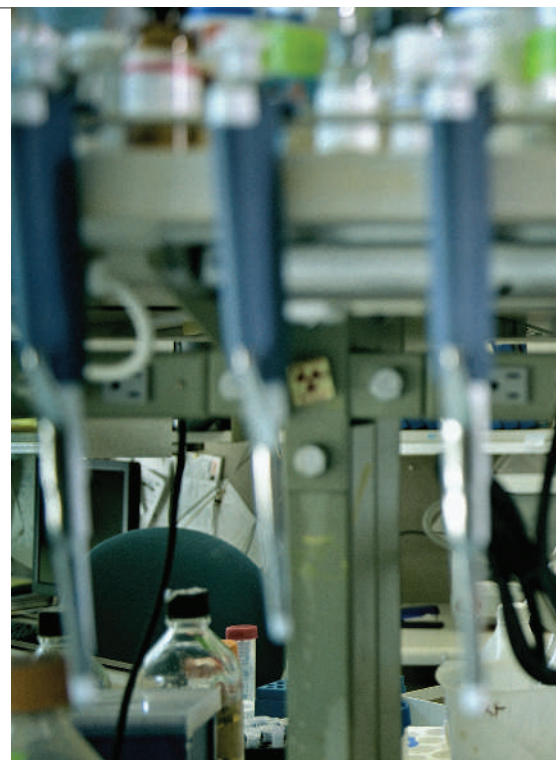
One of the most important factors is that risk assessment is now built into scientists' routines. Each chemical used comes with a list of potential risks and appropriate safety precautions, although unpredicted toxicity can affect even the most careful chemist, as Karen Wetterhahn found to her cost in 1996 (see 'Cautionary tales').

And practices such as eating lunch at the bench, mouth pipetting and washing hands with benzene (now known to be a carcinogen) have largely been consigned to history — apart from the odd emeritus professor reluctant to change methods that they have used for decades.

Better analytical techniques also mean that reactions can be run on much smaller scales, significantly reducing the danger, says Lawrence Gibbs, associate vice-provost for environmental health and safety at Stanford University, California.

"Academic institutions are now much more safety-conscious than when I was a student during the 1940s," agrees Edward Arnett, emeritus professor of chemistry at Duke University, Durham, North Carolina, and chair of the National Academy of Sciences research panel that penned the classic *Prudent Practices in the Laboratory: Handling and Disposal of Chemicals*. "I call that the kamikaze era of laboratory safety," he says.

"The thing that radicalized me on safety was

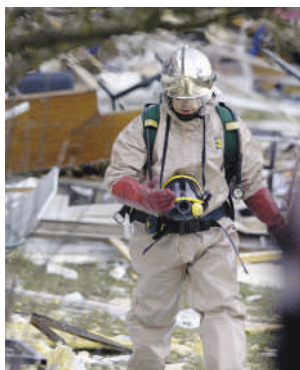


when I got a job with a small company in Philadelphia," he recalls. "I took over my job a couple of days after they buried my predecessor, who died of hydrofluoric acid poisoning. It was my job to make sure nothing like that ever happened again."

Yet despite the increase in legislation, it is surprisingly difficult to get national statistics on scientific accidents. In the United Kingdom and United States, for example, universities are obliged to report upwards only those accidents with serious consequences, such as hospitalization. This means that national agencies can remain oblivious to potentially major incidents if no one was seriously injured. And they do not tend to keep figures for different fields — the HSE, for example, groups all its accident figures for schools, colleges and universities into a single number, making it difficult to discern safety trends or to tell if one type of lab is more risky than another.

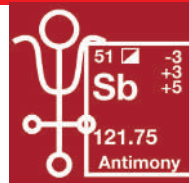
Having a national register of accidents would be "immensely useful," says Sara Cooper, director of health and safety at the University of Cambridge, UK. Because officials at the HSE are responsible for writing the codes of practice that university labs have to follow, "they need to understand, sector by sector, what we have to cope with," she argues.

It is possible to glean some information by talking to individual universities, however. Most accidents on campus involve trips and falls, like any workplace. Firn adds that most student injuries come from playing sports,



A fireman checks what's left of the Mulhouse laboratory.

B. FERRANDEZ/AFP/GETTY

**CHEMISTRY BLOG**

Read the Sceptical Chymist, *Nature's* forum for community chat about all things chemical.

<http://blogs.nature.com/the-scepticalchymist/>



Is anyone there?
Academic chemists
often end up working
alone late at night.

BRAD WILSON/GETTY

Cautionary tales

BLINDED BY ROOM TEMPERATURE

In 1970, K. Barry Sharpless was an assistant professor at the Massachusetts Institute of Technology. Having finished work for the day, and sans safety glasses, he checked in on a graduate student who was sealing a glass tube immersed in a bath of liquid nitrogen. Holding the tube up to the light, Sharpless saw the level of liquid in the tube drop suddenly — oxygen that had been inadvertently condensed by the chill of the bath was rapidly turning back into a gas. Before he could move, the tube exploded, and fragments of glass destroyed one of his eyes. In the following weeks, “the pain was terrific”, he remembers. While he retained vision in one eye, he says his lesson is straightforward: “There’s simply never an adequate excuse for not wearing safety glasses in the laboratory at all times.” Sharpless went on to win the 2001 Nobel Prize in Chemistry, and is now at the Scripps Research Institute in La Jolla, California.

UNEXPECTED POISON

When Karen Wetterhahn, a professor and toxic-metals expert at Dartmouth College, Hanover, New Hampshire, spilled a couple of drops of dimethyl mercury on her hand in August 1996, she quickly cleaned it up and assumed that her latex gloves had stopped the toxic chemical reaching her skin.

Five months after the accident, Wetterhahn was having difficulty walking and her speech started to slur. Tests later showed that she had 80 times the lethal dose of mercury in her blood. After losing her vision and hearing, she slipped into a coma and died in June 1997, aged 48. The tragedy shocked chemists, who were surprised at how easily the mercury had penetrated Wetterhahn’s glove and how toxic the compound was.

SAFETIES OFF

When the refrigerator-sized cylinder of liquid nitrogen exploded on 12 January 2006, it was standing on a tiled floor in a lab at Texas A&M University, College Station.

The blast peppered the lab’s walls and doors with tile shrapnel and the cylinder shot upwards, straight through the thick concrete ceiling and into the lab above, taking out windows, doors and an entire lab wall.

Investigators later found that two safety release valves on the cylinder had been removed and replaced with wire plugs, turning the cylinder into a pressure bomb. No one was hurt, but only because the explosion happened at about 3 a.m. **M.P.**

while stress is by far the major cause of ill-health in universities.

In labs, safety officers contacted by *Nature* agreed that chemistry generates the most accident reports. However, the vast majority of these involve undergraduates cutting themselves on glassware.

Ian Gillett, safety director at Imperial College London, cautions that such evidence may reveal little about the inherent dangers of chemistry. He points out that chemistry involves more practical work than other disciplines, chemists may be more vigilant about accidents than others, and in any case it is impossible to gauge how many minor disasters go unreported.

“I don’t feel that there is a significant difference between chemistry labs and other labs,” agrees Gibbs, pointing out that biology, for example, carries its own set of risks, namely infection.

Others agree that the risk of infection is often underplayed compared with that of chemical accidents. “People’s risk perception is skewed by the drama of an explosion,” says Firn. Gillett adds that the most serious lab accidents at Imperial have involved accidental infections with hepatitis A and vaccinia virus. “Pathogens are where I would be more anxious about what’s going on,” he says.

But what does seem clear is that academic labs are more dangerous than those in industry, with a more relaxed approach to safety.

“We find that the accident rate [in universities] is 10 to 50 times greater than in the chemical industry,” says James Kaufman, president of the Laboratory Safety Institute in Natick, Massachusetts. “In DuPont, if a guy hits his thumb with a hammer in Singapore, the chairman of the board has a report on his desk,” he says. “Imagine if that happened in academia.”

“In industry we often say that we are surprised more people aren’t injured in academic labs,” agrees Derek Lowe, a research chemist who blogs on “In the pipeline” (www.corante.com/pipeline).

“In universities, people are still learning, and people work all hours. If you are there alone at three in the morning, that’s seen as a good thing.”

Martin Pitt, compiler of *Bretherick’s Handbook of Reactive Chemical Hazards*, says that although risks have been much reduced in recent decades, there are still two key areas where academic labs could do a lot better. “Labs are too crowded,” he says, and such overcrowding raises the risk of spills.

Waste disposal is the other major issue. Most chemistry labs have open bottles where solvents are dumped along with the black gunk left from failed reactions. “It’s virtually impossible to say what will happen in these mixtures,” says Pitt. Such arrangements also make it too easy for people to dump noxious chemicals — and responsibility for them — in a communal spot. ■

“I took over my job a couple of days after they buried my predecessor.”

India boosts university places for lower castes

India has defied protestors and introduced legislation requiring universities and institutes of higher learning to reserve 49.5% of seats for students from lower castes. The change, to start in 2007, is a large increase from the current 22.5% quota.

Doctors have gone on strike against the move, and academics are critical. Students and researchers at the Indian Institutes of Technology and the Indian Institute of Science have appealed to President Abdul Kalam not to approve the legislation.

The government says it will spend US\$2 billion to increase the total number of students so that the same number of seats would be reserved for students on merit. But critics say this would dilute the standards of teaching.

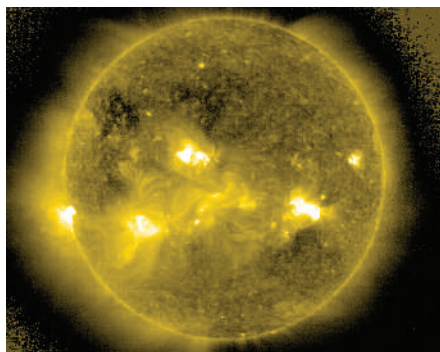
Solar satellite keeps on going and going...

It's been pronounced nearly dead, restored to health and temporarily blinded. Now the hardest-working satellite in solar physics, the Solar and Heliospheric Observatory

(SOHO), has had its mission extended until December 2009.

Launched in 1995 for a two-year mission, SOHO, a joint project of NASA and the European Space Agency, has survived several crises, including technicians losing contact with it in 1998, and solar storms in 2000 that blinded its instruments. But each time the satellite has pulled through.

SOHO will support new spacecraft studying the Sun, including NASA's twin STEREO satellites, due to launch in July, and the European Proba-2, slated for 2007. Project officials estimate that SOHO data have figured in more than 2,400 scientific papers published in peer-reviewed journals.



Staring at the Sun: the SOHO satellite, launched in 1995, has had its mission extended to 2009.

French engineer chosen as next leader of Caltech

Jean-Lou Chameau, a French-born civil engineer, will be filling David Baltimore's shoes come 1 September. Chameau will take over from the Nobel-winning biologist as the next president of the California Institute of Technology in Pasadena.

Chameau's background is in geoengineering. Most recently he served as provost at the Georgia Institute of Technology in Atlanta, another of the country's top engineering schools.

The choice is something of a surprise for the institute, whose past leaders include luminaries such as physicists Robert Millikan, also a Nobel laureate, and Harold Brown, the first scientist to serve as US defence secretary.

German society uses MP3 cash to build sound future

The MP3 digital audio format, developed in Germany ten years ago, has changed the music industry. Now it could also change the fortunes of innovative projects in applied research.

Germany's Fraunhofer Society, which

ESA/NASA

owns the patent on the MP3 technology, has announced that it will set up a foundation with the €100 million (US\$130 million) it earned last year from licence agreements. The idea has yet to be approved by the German government, but if it gets the go-ahead it will fund projects with market potential, such as the development of materials that can adapt to their environment.

The money would be used by the 58 Fraunhofer institutes throughout Germany.

Pakistani-Indian meeting axed after security checks

A conference between Pakistani and Indian scientists has been cancelled after the Pakistani government apparently retracted security clearances given to Indian scientists.

"This is a bizarre turn of events," says conference organizer Saleem Ali of the University of Vermont. The meeting was set for 29–31 May in Islamabad and would have brought Indian and Pakistani geologists together to discuss climatic and seismic issues facing both countries. It would also have explored the idea of turning the contested Siachen glacier into a scientific peace park (see *Nature* 440, 1; 2006).

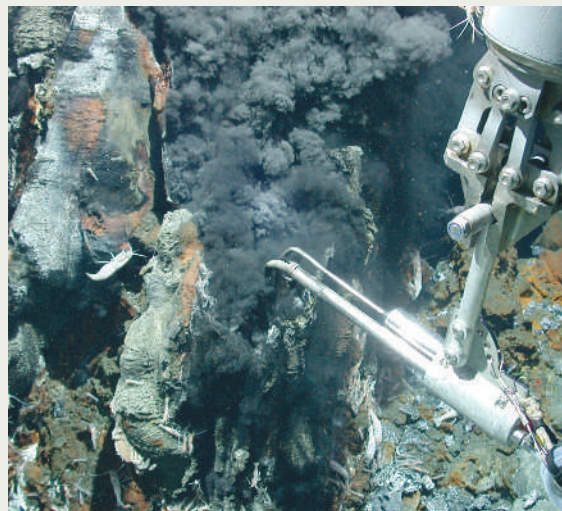
Ali says he is unsure why the clearances

Robot delves deep to find the hottest water

German scientists have measured the hottest temperature ever found at a hydrothermal vent — a crack in the ocean floor where superheated water pours out. Using a temperature sensor on a robot submersible (pictured), the team measured a scorching 407 °C at a depth of 3,000 metres in the equatorial Atlantic.

That's just 5 °C more than the previous record, which was measured in the Pacific. But it's enough of a difference to be important, as the pressure and temperature are great enough to turn the water into a supercritical fluid — a sort of fluid-gas hybrid state. The researchers hope to learn more about the elements dissolved in the mixture.

The team, led by Andrea Koschinsky of the International University of Bremen, Germany, is continuing to probe southwards along the underwater Mid-Atlantic Ridge.



MARUM, UNIV. BREMEN

were revoked. "The Pakistani military seems to have got involved," he says, adding that military officials visited several local organizers. He doubts that the conference can now be rescheduled.

Correction

There was an editing error in our News story on scientific misconduct (*Nature* 441, 392–393; 2006). We should have said that Jin Chen was at Shanghai Jiaotong University.



Tropical storms, such as Cyclone Larry, pictured here bearing down on the coast of Australia, may become more intense in a warmer world.

Bad weather ahead

Scientists and policymakers are battling over whether global warming is making hurricanes more destructive.

Alexandra Witze ventures into the heart of the storm.

It remains the worst tempest in written history. In November 1970, a tropical storm strengthened into a cyclone in the Bay of Bengal. It slammed into what is now Bangladesh, killing up to half a million people. But just try digging up specific data about it. You won't find it. In the database of Indian Ocean storms, the wind speeds for this particular cyclone are listed as zero.

For storm researchers, this is a wearily familiar tale. The historical data on tropical cyclones are notoriously patchy. Some gaps are down to irregularities in the measurements collected during past storms. Others are caused by plain old human error. The Bangladesh cyclone, for example, happened during a period when databases from two government agencies were being merged and data were lost in the transition.

Today, increasingly sophisticated satellite imagery makes gaps less likely. But researchers still wrestle with the imperfect historical record — especially if they are trying to understand whether global warming is making

cyclones more intense. Some argue that such a claim is hard to prove if major storms of the past aren't on record.

As the 2006 hurricane season gets under way in the Atlantic basin, few issues could be hotter than the relationship between global warming and tropical storms. Forecasters predict that things won't be as bad as 2005, which saw a record 28 named storms in the Atlantic and probably more than US\$100 billion in damages (see "The season ahead"). But authorities are looking to scientists to tell them whether 2005 is an example of a hurricane season that we will have to get used to.

At first glance, a link between cyclones and global warming seems to make sense. Tropical cyclones are born over the oceans, where masses of rotating air pick up ever more energy from warm surface water. Once the winds in the mass reach 33 metres per second, a tropical cyclone is born. In the northwest Pacific, it's called a typhoon; in the Atlantic and northeast Pacific, a hurricane; elsewhere, a cyclone.

NASA

But only recently have scientists come up with the data that suggest global warming makes cyclones more intense. Two major studies laid the groundwork last year. In the first, published in August, atmospheric scientist Kerry Emanuel proposed that hurricanes had grown more intense over the past 30 years, most likely because of increasing sea surface temperatures¹. Emanuel, of the Massachusetts Institute of Technology, developed an index to describe how destructive a storm could be, and found that the wrecking power of storms correlated strongly with sea surface temperature.

Tempests of doom

The second paper² came in September, soon after Hurricane Katrina had killed more than 1,800 people along the US Gulf Coast. A team led by Peter Webster, of the Georgia Institute of Technology in Atlanta, studied the occurrence of storms rated at the higher end of a strength-categorization scale called the Saffir-Simpson scale.

Hurricanes are ranked from 1 to 5 on the scale: storms with wind speeds reaching 33 metres per second are at the low end of category 1, and the threshold wind speed for a category 5 storm is 67 metres per second (or 241 kilometres per hour). Hurricane Katrina was category 5 when over the Gulf of Mexico, and had weakened to category 3 when it slammed into the Gulf Coast. Webster's team reported that there has been a rise in the number of category 4 and 5 storms in the past 35 years, in nearly all of the world's ocean basins².

Together, the Emanuel and Webster papers kick-started fresh efforts in a previously obscure corner of meteorology. A veritable flood of findings has emerged; some preliminary work was presented in April at a meteorology meeting in Monterey, California³. "We've moved forward immensely since last June," says Greg

Holland, a co-author on the Webster paper and a meteorologist at the National Center for Atmospheric Research in Boulder, Colorado.

Yet many research areas remain untapped. One major unknown, experts say, is how hurricanes interact with the ocean — not just form above it. "Right now, almost everyone attacks the problem as hurricanes responding passively to climate change," says Emanuel. "They are active players." Hurricanes leave a trail of cold water in their wake — which is not currently accounted for in most climate models.

Other researchers point to the need to better understand the factors affecting hurricane intensities. They hope that data from last year's 'hurricane-hunter' flights will help⁴; these involved forecasters flying into the heart of Atlantic hurricanes to find out what drives changes to their intensity. And mysteries still surround the issue of how hurricanes form in the first place. Indeed, one speaker, David Nolan of the University of Miami, Florida, drew a crowd in Monterey with his provocatively titled talk, 'Could hurricanes form from random convection in a warmer world?' (His answer: 'no').

Given that more research is obviously needed, how should scientists best direct their efforts to get useful answers as soon as possible? Many echo the adage that the past is the key to the present, and argue that far more time and money need to be spent on paleotempestology — the study of past hurricanes, as recorded in geological deposits. Studies that look back as far as several thousand years ago could help resolve the frequency with which hurricanes form, or at least make landfall in certain regions of the world. These would provide an invaluable measure of 'normal' patterns against which to weigh modern trends.

Perhaps most crucially, meteorologists say,

the renewed interest in hurricanes could inspire researchers to work on improving the historical record of storms. It is possible to go back through the database and re-assess each storm with modern eyes, making sure its strength and trajectory are analysed by the same standard as more recent storms. That's what Christopher Landsea, a meteorologist at the National Hurricane Center in Miami, has been doing for the Atlantic hurricane database, which contains measurements on storms dating back to 1850.

Elusive evidence

Re-analysing past measurements is one thing. But the biggest problem, says Landsea, is in having to work with a lopsided data set. If you knew what was in sausages, you wouldn't want to eat them, he says; likewise the historical record shouldn't be trusted. For instance, Hurricane Wilma garnered headlines last summer

when it was recorded to have the lowest central pressure — another measure of storm intensity — of any known hurricane in the Atlantic basin. Yet Wilma "was

sampled just about every hour of its existence", says Landsea. Compare that, he says, to a tropical storm such as Carol, which moved up the US eastern seaboard for days in 1954 but was sampled only seven times over its lifetime.

The picture gets even bleaker in the world's other ocean basins. In a recent informal study, Landsea looked through satellite images of storms in the northern Indian Ocean. From these pictures, he estimated that the storms should have been rated as category 4 or 5; they were recorded as being of lower intensities at the time. Landsea says that if these storms are missing from the records, how is it possible to conclude that hurricane intensities are increasing because of global warming? They could be,

"The name-calling has got to stop."

— Max Mayfield



Entire plantations were wiped out when Cyclone Larry (left) struck land earlier this year.

T. BLACKWOOD/AFP/GETTY

BETTANN/CORBIS



Destructive force: a cyclone that hit what is now Bangladesh in 1970 killed up to half a million people.

he says; it's just impossible to tell.

Webster disagrees: "Chris has found several category 4s and 5s we missed in the early 1970s; he has to find 152 for us to be wrong." And Emanuel adds that, although one can argue about the particular number of hurricanes in a particular year, his measure of hurricane intensity still correlates strongly with sea surface temperature — no matter how many storms there are in a particular year.

Divided opinion

The disagreement echoes deep battle lines between several camps, many of which have been re-ignited by the recent studies. A flurry of critiques has appeared in *Science* and *Nature*, as well as in the blogosphere. The debate has got personal at times, and few are

happy about it. "The name-calling has got to stop," says Max Mayfield, director of the National Hurricane Center in Miami.

In one recent paper, longtime climate-change sceptic Patrick Michaels and colleagues argue that rising sea surface temperature no longer affects the intensity of a hurricane once its winds have reached speeds of more than 50 metres per second⁵.

In another, Philip Klotzbach of Colorado State University in Fort Collins writes that there is no strong correlation between hurricane energy and sea surface temperature in most of the world's ocean basins — and that Webster's and Emanuel's results are due mostly to the patchiness of data sets prior to the mid-1980s (ref. 6). And the *Bulletin of the American Meteorological Society* has hosted a feisty back-and-forth, spearheaded by policy expert Roger Pielke Jr of the University of Colorado. He calls links between hurricanes and global warming "premature".

For many, the stakes could not be higher. Knowing where and how often storms might strike is crucial for shaping government policies. Exploding populations in coastal zones place ever-greater numbers of people at risk — a fact noted by some policy experts, who say that the apparent increase in hurricane destructiveness seen in the past few years is down to the fact that more people are living in at-risk areas⁷.

Preliminary studies by other groups seem to

bear out Webster's and Emanuel's conclusions. A new study of Indian Ocean hurricanes, presented at the Monterey meeting, suggests that there has indeed been an increase in category 4 and 5 storms in the region — and few Indian Ocean storms are missing from the database. And using a data set of global storms that occurred between 1958 to 2001, scientists from Purdue University in West Lafayette, Indiana, have found the same overall increase in storm destructiveness in recent years — particularly after 1985 (ref. 8).

Worse to come

Other scientists are turning to computer models for possible answers to questions, such as how much will sea surface temperature rise, and exactly how will that influence hurricane formation? Computer models suggest that sea surface temperatures in the Atlantic hurricane-forming region could warm by 2 °C by 2100. They also suggest that if this rise occurs,

maximum wind speeds could increase by 6% (ref. 9). It may not sound like much, but damage from hurricanes rises in proportion to the cube of the wind speed.

So far, the world's oceans haven't seen anything close to a 2 °C warming — just a 0.5 °C rise since 1970. "The warming we've seen to date is really just the tip of the iceberg," says Thomas Knutson, a climate modeller at the Geophysical Fluid Dynamics Laboratory in Princeton, New Jersey.

Predicting future hurricane activity will also require greater understanding of how natural climate fluctuations interact with global warming. For example, the El Niño Southern Oscillation, a pattern of temperature fluctuations in the tropical Pacific ocean, can affect the formation of hurricanes in certain regions, as can volcanic eruptions.

For people living in vulnerable coastal regions, answers to the debates can't come soon enough. In the United States, last year's destruction has prompted a new round of calls for hurricane preparedness and mitigation measures. New Orleans, the city devastated by Katrina, is slowly being rebuilt with higher levees to keep storm surges out. And while researchers argue over the details of databases and data analysis, residents are girding for another six months of uncertainty.

Alexandra Witze is *Nature's* chief of correspondents for America.

"The warming we've seen to date is just the tip of the iceberg."
— Thomas Knutson

THE SEASON AHEAD

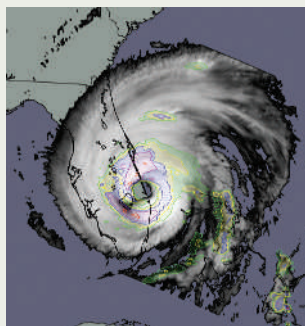
Given all the attention to Atlantic Ocean hurricanes, it's hard to remember that only 11% of the world's tropical cyclones occur there.

In 2005, that attention seemed warranted. Seven hurricanes made landfall in the United States. Katrina killed 1,800 people. There were so many storms that the National Hurricane Center ran out of alphabetized names and started giving them Greek names instead.

Experts say the United States should prepare for more of the same. The latest predictions suggest that sea surface temperatures in the region where hurricanes are born are running higher than average. The National Oceanic and Atmospheric Administration predicts that there will be up to 16 named storms in the Atlantic this season, with up to ten

becoming hurricanes. William Gray, a forecaster at Colorado State University, and his colleagues have predicted 17 named storms, with nine of them hurricanes.

Not everyone agrees on the significance of such



numbers. Gray, for his part, is one of the most vocal critics of a possible link between hurricanes and global warming. He argues that the Atlantic is seeing nothing more than a peak in a regular cycle of activity. The 1950s

and 1960s were active decades; in the 1970s, things quieted down. Activity picked up again in 1995 and has increased ever since.

Others disagree with Gray. "We may have quiet years, but right now it doesn't look like a quiet decade," says Kerry Emanuel, a hurricane expert at the Massachusetts Institute of Technology.

For this year, coastal residents should pay close attention to what happens in the next two months. Roughly 90% of the Atlantic's hurricane activity takes place after 1 August, says Eric Blake of the National Hurricane Center in Miami. Adding sea surface temperature data from June and July should radically improve forecasts for this coming season, he says — letting people know how much 2006 will look like 2005.

A.W.

NASA

1. Emanuel, K. *Nature* **436**, 686–688 (2005).
2. Webster, P. J., Holland, G. J., Curry, J. A. & Chang H.-R. *Science* **309**, 1844–1846 (2005).
3. Witze, A. *Nature* **441**, 11 (2006).
4. Schrope, M. *Nature* **438**, 21–22 (2005).
5. Michaels, P. J., Knappenberger, P. C. & Davis, R. E. *Geophys. Res. Lett.* **33**, doi:10.1029/2006GL025757 (2006).
6. Klotzbach, P. J. *Geophys. Res. Lett.* doi:10.1029/2006GL025881 (2006).
7. Pielke, R. A. Jr, Landsea, C., Mayfield, M., Laver, J. & Pasch, R. *Bull. Am. Meteorol. Soc.* **86**, 1571–1575 (2005).
8. Srivier, R. & Huber, M. *Geophys. Res. Lett.* (in the press).
9. Knutson, T. R. & Tuleya, R. E. *J. Climate* **17**, 3477–3495 (2004).

Cashing in on the rich coast

Costa Rica's flagship conservation institute needs help. Can a new deal with industry save it? **Rex Dalton** investigates.

Along Costa Rica's Pacific coast, scientific explorers are trying to turn over a new leaf for a storied institute — the National Biodiversity Institute, or INBio.

Near a crocodile-infested river in Guanacaste province, an international team searches for bacteria that one day may become a drug or industrial product. Taking advantage of the dry months, they are focusing on a microbe-rich forest zone — hoping to flip a leaf to find a fungus, either attacker or defender¹. If they get lucky and isolate a fungus that produces a useful compound, any economic rewards will be shared with INBio, a once-model organization that is struggling to survive.

Created in 1989, INBio, based in a suburb of the capital San Jose, became an early symbol for how developing nations might participate sustainably in the biotechnology revolution. World-class researchers joined with Costa Rica's well-trained academics, hoping to save the nation's biodiversity — 4% of the world's total — by making money from it.

In the first years, more than US\$4 million in grants flowed from the drug giant Merck². Foundations and other nations added more, with total donations eventually topping \$63 million. But no major products emerged.

There have been some modest successes, including a couple of industrial compounds and an over-the-counter hangover remedy. More significantly, INBio set itself up as a training ground for marrying traditional conservation values with modern high-tech methods. Today, other developing nations look to INBio as an example of how to achieve the goals of the 1992 Convention on Biological Diversity, which encourages sustainable development worldwide.

But INBio's scanty returns, and Costa Rica's limited ability to fund research, has raised doubts about the institute's survival. Its future may depend on an ambitious new bioprospecting plan, funded by the United States, in which scientists will canvas the country for new drug candidates. Others are holding out hope for a far-reaching goal for a \$500 million endowment to preserve a quarter of Costa Rica's biodiversity (see 'A fund for the future').

"INBio is a high-quality machine with no gasoline," says Daniel Janzen, an entomologist



INBio has a huge collection of Costa Rican insects, but exploiting their chemistry has proved tough.

at the University of Pennsylvania in Philadelphia who has worked in the country for decades and is trying to create the endowment. "The sooner we can get gas in the system, the sooner it can be cranked up."

Fuel source

Advocates hope the fuel for that engine will come from the Fogarty International Center at the US National Institutes of Health, which is giving the new bioprospecting team \$3.5 million over four years. The effort is one of several International Cooperative Biodiversity Group (ICBG) projects, designed in part to put a value on preserving biodiversity³.

Led by chemist Jon Clardy of Harvard Medical School, the five-year project includes researchers from the University of Michigan in

Ann Arbor, the Broad Institute — a joint venture of Harvard and the Massachusetts Institute of Technology — and the Novartis Institutes for Biomedical Research in Cambridge, Massachusetts. Team members will be happy with whatever they find, but their assays and screens are designed to find compounds that can fight malignancies, infectious agents or neurodegenerative disease.

For five years ending in 1997, INBio was involved with a different ICBG, led by researchers from Cornell University in Ithaca, New York. But that programme produced no money-making products, and, in hindsight, some scientists believe its goals — such as trying to extract useful compounds from insects — were unrealistic.

The current project is designed to learn

A fund for the future

"Nibble by nibble," Costa Rica's fabulous biodiversity is disappearing, says entomologist Daniel Janzen.

The way to save precious habitats, he says, is to raise a US\$500 million endowment. Annual earnings then would fund conservation efforts — hopefully saving 25% of the country's land. "Right now, it is death by a thousand cuts," says Janzen, of the University of Pennsylvania.

In recent months, he and colleagues have prepared the concept to present to Oscar

Arias, recently reinstalled as the country's president. They call it Costa Rica Sostenible — sustainable Costa Rica.

On 26 May, the researchers were to brief ministry officials on the endowment concept. On 2 June, key government and environmental players will meet to begin planning the process to create it. "I've come to the conclusion that biodiversity won't survive if we don't do something like this," says Janzen.

He bases this on 40 years of studying some of the 9,600 species of butterflies, moths and

caterpillars in the Pacific coastal province of Guanacaste — equal to the number in the United States and Canada combined.

Alvaro Ulgalde, a conservation biologist in San Jose who created Costa Rica's national park system 36 years ago, endorses the endowment idea. "It makes sense in every respect," he says. "But it needs to be complemented by clear government enforcement policies."

Under the plan, about \$100 million would be used to buy pockets of private land within the country's national parks and

other reserves. The remaining \$400 million would be invested outside Costa Rica.

The estimated \$20 million that would result annually would be shared equally by the nation's 11 conservation areas, with the National Biodiversity Institute (INBio) also receiving an equal portion for its conservation efforts.

Powerbrokers in San Jose are enthusiastic, particularly as the money will be outside the country and well accounted for, an aspect they hope will appeal to donors.

R.D.

from the past. In particular, it is structured to be as open as possible about any potential new drug candidates. With the earlier grants, any compounds of interest left Costa Rica, disappearing behind the proprietary walls of corporate science. Clardy, who was part of the earlier ICBG when at Cornell, says he didn't want the new programme making the same mistakes. "I specifically formed it this way," he says.

To Rodrigo Gamez — a Costa Rican plant virologist who was key to forming INBio and remains president of its governing body — the experience was invaluable in securing the new pact. "We developed a capacity to negotiate," he says.

But some question how hard the companies may or may not have tried to develop products in those earlier deals. Merck, for instance, is believed to have found that about 200 substances tested positively in preliminary screens against disease. But Clardy's group thinks the firm did not go far beyond initial screening.

"I think they bought good public relations with the grants," says one informed scientist, privately. Merck declined to respond. Similarly, in a separate deal, Bristol-Myers Squibb is believed not to have pursued promising signals from compounds it may have found in insects. Neither company has said much publicly about results from the tested compounds.

Bugged out

In the earlier ICBG, scientists from Cornell and INBio set out to learn from the country's insects, which use an array of chemicals for digestion and protection. Insects were a natural choice, given INBio's extensive documentation of the country's butterflies, moths and caterpillars.

But the team found insect substances were present only in incredibly small quantities, making the material scarce for studies. And culturing them to produce enough material for lab experiments was onerous or worse.

Clardy eventually came to believe that working with insect material was too difficult. "We are only going to work on compounds that can be easily cultured and duplicated for studies" — from fungi, leaves, or other sources, he says.

In addition, the data will be publicly accessible, in a database containing information such as where the compounds were collected and under what conditions. Clardy foresees an eventual library with some 5,000 to 10,000 compounds collected during the project. The database could even contain details on how compounds respond in various screening tests against pathogens, information that is usually considered proprietary.

Clardy's group would get first shot at studying any promising disease-fighting compounds. But eventually the data would enter the publicly accessible ChemBank.

The work is possible because academic groups, such as those at the Broad Institute



Hot property: the rivers of Palo Verde National Park, in northwest Costa Rica, could be a source of useful bacteria.

and Harvard's Institute for Chemistry and Cell Biology, are getting hold of the sophisticated and expensive testing equipment that in the early 1990s, when the first projects got under way, was the preserve of drug companies. In Novartis, the project has a partner willing to adapt to the new environment — and one that apparently still has faith in drug development from natural products, a faith that many pharmaceutical companies have lost.

As a team member, Novartis will get first opportunity to run its proprietary assays on promising compounds. But then the firm will have to negotiate agreements with INBio to advance any material to drug stage.

"I am very comfortable with this set up," says chemist Alexander Wood, a Novartis executive director for oncology research. "We want to have as many opportunities from new compounds as possible. When it comes to design, medicinal chemists can't match the natural process."

Revenue trickle

Such a deal might also address the lingering issue of a developing nation sharing in benefits from bioprospecting — which some activists call biopiracy.

Historically, countries housing microbes that led to blockbuster drugs got virtually nothing. The diabetes drug acarbose, for instance, was derived from a bacterium in a Kenya lake; proceeds from its sales, about \$380 million in 2004, go entirely to the drug company Bayer, which developed it.

Under the diversity convention, a nation such as Costa Rica — which has adopted the pact — is to secure some reward from any pharmaceutical proceeds. In the past few years, Costa Rica has adopted and implemented laws specifically to address this access and benefit sharing.

Costa Rica — and in turn INBio — is among the few developing nations currently receiving license fees from natural products. Diversa, a San Diego-based biotech company, is currently paying Costa Rica nearly \$6,000 a year for two products developed from the country's resources. One is DiscoveryPoint, a fluorescent protein used for tagging material in experiments that comes from a marine organism found in the Caribbean Sea. The second is Cottonase, an enzyme for processing raw textile material to reduce the use of harsh chemicals. It was discovered in warm mud in a volcanic area just west of INBio's suburban campus.

And INBio's knowledge of plants helped a Costa Rican firm, Lisanatura, develop a treatment for hangovers or indigestion — Hombre Grande — from a plant called amargo (*Quassia amara*).

Diversa is also persisting with the insect



Gut instinct: INBio staff inspect a termite mound. US collaborators are probing insects' enzymes and symbionts.

world. Along with the California Institute of Technology in Pasadena and the US Joint Genome Institute in Walnut Creek, California, the company has contracted with INBio staff to probe the guts of Costa Rican termites for useful compounds. The project involves analysing how termites use microorganisms or enzymes to dissolve cellulose. Some termites have 100 species living in their gut⁴. The team

"With the earlier grants, any compounds found left Costa Rica, disappearing behind the proprietary walls of corporate science."

isolates organisms or enzymes, sequences key genetic material, and then studies it.

INBio's annual operating budget is around US\$6 million. About 70% of that comes from grants and contracts, such as those from the ICBG and Diversa. In addition to bioprospecting, INBio has two other divisions — a nature park, where schoolchildren and others learn about the importance of preserving biodiversity, and its inventory work, largely involving Janzen. With such revenue sources, when grants run out it can mean staff layoffs; late last year, about 15 workers were let go.

Those who remain are dedicated. In March, project leaders laid out an ambitious schedule for collecting, culturing and screening compounds. In Palo Verde National Park, they began searching river channels and an estuary

on the Gulf of Nicoya. Scooping up bottom silt, they immediately found hair-thin, noodle-like filaments that are a type of cyanobacteria.

So far, no drug has come from cyanobacteria. Their history traces back more than two billion years, and today there are hundreds of types that thrive in salt, fresh and brackish waters. Such bacteria are known to host protective molecules, which chemists hope to isolate⁵.

Organic chemist David Sherman of the University of Michigan has joined with Jorge Cortes, a University of Costa Rica marine biologist who is an authority on sea fans. Since 2000, Sherman has been diving in oceans from Papua New Guinea in the South Pacific to waters off the Central American isthmus in search of new compounds⁶. Together, he and Cortes explore the environments cyanobacteria thrive in.

Local knowledge

In December, as part of an ICBG workshop, the pair found cyanobacteria while diving on reefs in the Pacific Gulf of Santa Elena, just south of Nicaragua. "The diving is very challenging — low visibility, a pronounced surge, and high surf," says Sherman. "It was really great to have a local expert involved."

In the search for fungi on land, the process also involves unravelling the role of microbes. On a field trip this spring, INBio chemist Giselle Tamayo noted areas of interest in trees and shrubs. Two metres up from the ground, the vegetation is rich with fungi. Leaves will be collected and treated with antibiotics at INBio's labs. Then the fungi-bearing material will be cultured for weeks.

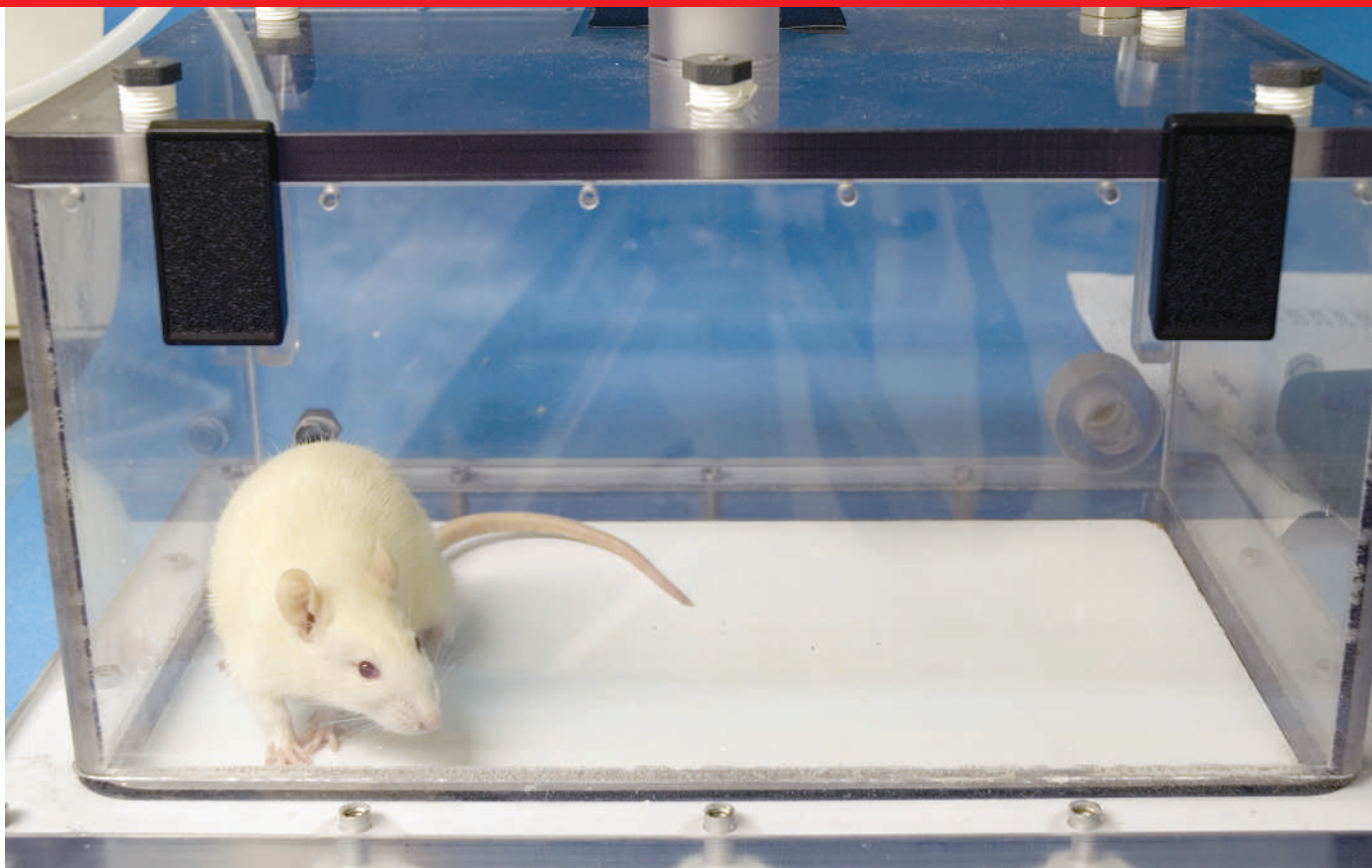
Periodically, fungi will be removed from the sample vials. "The last ones to grow are what we are interested in," says Tamayo. "These slower growing ones probably have not been described before."

Already, the team is seeing they will need luck. In the laboratory, technicians isolated DNA from water in a Costa Rican bromeliad plant. The DNA segments were then inserted into *Escherichia coli* bacteria, which produced a compound that has shown antibiotic activity in culture tests⁷. "We did get lucky with that," says Clardy. "But it could be nothing, and it could fall apart tomorrow."

The same, they hope, won't happen to INBio.

Rex Dalton is Nature's West Coast correspondent.

1. Clardy, J. & Walsh, C. *Nature* **432**, 829–837 (2004).
2. Aldhous, P. *Nature* **353**, 290 (1991).
3. Weiss, C. & Eisner, T. *Technol. Soc.* **20**, 481–498 (1998).
4. Brennan, Y. et al. *Appl. Envir. Microbiol.* **70**, 3609–3617 (2004).
5. Fortman, J. L. & Sherman, D. H. *ChemBioChem* **6**, 1–19 (2005).
6. Salomon, C. E. et al. *Nat. Prod. Rep.* **21**, 105–121 (2004).
7. Brady, S. F. & Clardy, J. *J. Nat. Prod.* **67**, 1283–1286 (2004).



H. GOLLEDGE

An easy way out?

Scientists say they gas mice and rats with carbon dioxide because it is humane. It's also simple, cheap and keeps their hands clean. **Emma Marris** analyses the final seconds of the lab rodents' life.

"One uses CO₂ to knock the animal out, and it takes a bit longer than you would like — about a minute," explains Abigail Witherden, a kidney researcher at Imperial College London describing the death of a mouse. Then, to be sure: "you break its neck".

This is Witherden's preferred method of execution. "You feel more brutal if you've got a wriggling, living creature and you break its neck," she says, "rather than something that is probably already dead. It is worthwhile considering that some of these things are done for the benefit of the people, not the mouse."

Somewhere, right now, a lab mouse or rat is meeting its end, almost certainly by being gassed with carbon dioxide. The gas has become the preferred method because it is easy, cheap, and untroubling for the researchers. The rodent appears to drift off to a peaceful death. But some scientists say that CO₂ could cause pain and distress, and that it might be better to use a pricier anaesthetic, or a manual manoeuvre called cervical dislocation that breaks the animal's neck.

Tens or perhaps hundreds of millions of laboratory mice and rats are killed each year. The exact number is unknown, as there is no

reporting requirement in the United States, probably the heaviest user. In the United Kingdom, roughly 2.4 million rodents were used in experiments in 2004, according to the Home Office, which oversees lab animals¹.

Big sleep

Their use is on the rise. "While researchers' needs for animal models are declining, the use of rodents is increasing because of transgenic models — knockout mice and so forth," says George Goodno, spokesman for the Foundation for Biological Research, a Washington DC-based lobby group for animal research. Scientists can now buy transgenic rodents to

order, and their commercial production involves culling those that have the wrong genes, are unhealthy or have become surplus inventory.

In typical CO₂ gassing, one or more rodents are placed in a chamber and CO₂ is gradually added. As the CO₂ level rises, it acts as an anaesthetic and eventually knocks the rodent out. Then the lack of oxygen kills it. Adding CO₂ slowly is intended to avoid the pain caused by high concentrations, when the gas dissolves on the mucous membranes and forms carbonic acid. The hope is that the animals are out cold by this stage.

In the United States, rodents are sometimes put into chambers that have been pre-filled with CO₂, which kills them within 40 seconds, but might be more painful. The United Kingdom bans this method. None of these methods is foolproof. The Office of Lab Animal Welfare at the US National Institutes of Health (NIH) in Bethesda, Maryland, gets occasional reports of animals that wake up after being removed from the chamber. It is to avoid this that many labs, including Witherden's, break the gassed animals' necks.

There are rules about how to kill a mouse or rat, but many of the details are left up to the lab. The major guidance documents on killing



"There is a strong suspicion that carbon dioxide levels that in no way cause rodents pain do cause them distress."
— Huw Golledge

H. GOLLEDGE

methods for researchers in the United States, Europe and Japan are all being revised; the use of CO₂ is likely to be a point of dispute.

In the United States, researchers using federal funds must follow the 2000 report of the American Veterinary Medical Association Panel on Euthanasia². This report, though, is less of a how-to guide than a summary of the pros and cons of various methods. A few are deemed unacceptable; others, including gassing with CO₂ alone, are ruled as acceptable, because they consistently result in a humane death. Cervical dislocation is ruled as “conditionally acceptable”, partly because it takes some skill. The report also considers the effects of various methods on the researchers who kill the animals — but also says that “there are occasions, however, when what is perceived as aesthetic and what is most humane are in conflict”.

This February, a group of 34 scientists and those in the lab-animal business gathered at the University of Newcastle upon Tyne, UK, to discuss the possibility that CO₂ is not the most humane killing method. Huw Golledge, a neurophysiologist at Newcastle who helped to organize the conference, says that his work has convinced him that, at least with rats, loss of consciousness can be assured before the CO₂ reaches painful levels, if one adds the gas at less than 30% per minute. But before they pass out, are the rodents gasping for breath and panicking?

Breathless

“There is a strong suspicion that CO₂ levels that in no way cause them pain do cause them distress,” says Golledge. Given the choice, rats will rapidly get out of a chamber containing CO₂ concentrations of about 15% or more³ — far below the concentration needed to knock them senseless.

What convinced Golledge was the work of Robert Banzett at Harvard University, who studies the distress caused by breathlessness, or dyspnoea, the sensation that one cannot breathe properly. “He has tried to make animal models for his human research,” says Golledge, “so we got him over. My opinion on CO₂ until this meeting was that if you did it right, it was okay. Having heard what Bob Banzett said about dyspnoea, I think we need to clear that up.”

Kathleen Conlee of the Humane Society, an animal-welfare organization based in Washington DC, doesn't need any more evidence. She thinks killing rodents with CO₂ alone is cruel, and prefers an anaesthetic gas — or, in cases where the anaesthetic would interfere with post-mortem measurements, the guillotine⁴. “We say, if there is a small number of animals, and the person is well trained, use decapitation. If you are doing a large number, use halothane or isoflurane, and then blast them with however much CO₂ you want.”

“I feel strongly about it personally,” Conlee adds. “The number of animals is huge; the

evidence is here; the alternatives are here. It is a no-brainer.”

But moves away from CO₂ face two barriers: researchers' comfort levels and economics. “It is so widely used, and so convenient, and people have used it for so long, that it is going to be very difficult to change it,” says David Morton, head of the centre for biomedical ethics at the University of Birmingham, UK, who helped to write a scientific report for the European Commission on the topic. The commission will use the report when revising its directive on lab-animal welfare⁵.



“I feel strongly about it personally. The number of animals is huge; the alternatives are here. It is a no-brainer.”
Kathleen Conlee

The report argued for the use of anaesthesia before CO₂, saying killing with CO₂ alone “should be phased out as soon as possible”. But in large operations, setting up facilities for an extra anaesthetic would involve extra costs and regulations. Still, Morton believes it is worth pushing for: “The ethical mandate for me is that if you can show there is a welfare problem, then you have to do something about it.”

Other researchers, though, feel that the literature on CO₂ is too limited and contradictory to start moving away from it. “There are differences of opinion on the use of CO₂, and much of this stems from the inconsistency in the literature. There is very little consistency of



Hands-on: some researchers advocate cervical dislocation to ensure that gassed mice are dead.

methodology in these papers,” says B. Taylor Bennett, associate vice-chancellor for research resources at the University of Chicago. “We need research done, but there are very few sources of funds for this type of research.”

A good death?

Meanwhile, many researchers remain comfortable with using CO₂ alone. “If it is done properly, I don't have any problems at all using CO₂,” says cell biologist Claire Pollock, who, when she began killing mice at the NIH's National Cancer Institute in Bethesda, allayed her fears by researching the science on the subject and throwing herself into her training courses. “I can't stress enough that any technique is only as good as the researcher,” she adds. “I would much rather be put into a CO₂ chamber than be put into the hands of someone who doesn't know what they are doing for cervical dislocation.”

Indeed, many researchers use their own intuition to decide what makes a humane death, perhaps because lab rodents can't tell us what they are feeling. And even in aversion studies, rodent behaviour can be hard to interpret. Golledge, for example, says he would prefer slowly rising CO₂ over a dive into a pre-filled chamber. Robert Banzett reportedly says he would make the opposite choice. Likewise, alternatives to CO₂, such as carbon monoxide, are discussed with analogy to humans. Inhaling carbon monoxide is a common method of suicide, largely because it is said to be painless. “And,” points out Golledge, “you hear about people who have a malfunctioning gas boiler, and they just fall asleep [and die] in their armchairs.” Another possibility is argon gas, to which lab rats seem less averse³. Argon is popular in the poultry industry and wins support even from animal-welfare campaigners. “It should be looked into,” says Conlee. “Sometimes we hear about concerns with the anaesthetic gas [affecting] personnel, and that wouldn't be a concern with argon.”

In any case, all of these lab techniques are undoubtedly more humane than most domestic methods of killing unwanted rodents, from snap traps that sometimes merely paralyze, to sticky traps that capture the mouse, requiring some kind of blunt instrument or drowning. “No matter what you do, it is going to be more humane than at home,” says Witherden. ■

Emma Marris reports for Nature from Washington DC.

1. Home Office Statistics of Scientific Procedures on Living Animals: Great Britain 2004 (HMSO, London, 2005). www.official-documents.co.uk/document/cm67/6713/6713.pdf
2. AVMA Panel on Euthanasia *J. Am. Vet. Med. Assoc.* **218**, 669–696 (2001).
3. Leach, M. C. *et al. Comp. Med.* **52**, 249–257 (2002).
4. Conlee, K. M., Stephens, M. L., Rowan, A. N. & King, L. A. *Lab. Anim.* **39**, 137–161 (2005).
5. European Food Safety Authority AHAW Panel Aspects of the biology and welfare of animals used for experimental and other scientific purposes. (2005). europa.eu.int/comm/environment/chemicals/lab_animals/pdf/efsa_scientificreport_en.pdf

BUSINESS

Upending the zeitgeist

Germany's best-known stem-cell researcher is leading a charge to build up more commercial acumen on the nation's university campuses. Ned Stafford reports.

The biggest experiment at Life & Brain, a new campus-based biotechnology company at Bonn Medical School, is Life & Brain itself.

Will it become a commercial success within German academia or — as some conservative professors might hope — will it fail? Oliver Brüstle, company co-founder and head of reconstructive neurobiology at the university, believes the project, housed in a chic five-story building, will be vindicated. And he wants it to send a message to the rest of German academia: the relationship between universities and commerce must change.

The company doesn't yet have any products, but it does have millions of euros of backing from the state government and from the university. Brüstle, who gained prominence in Germany five years ago as a leading public advocate of human embryonic stem-cell research, wants to apply stem-cell research commercially. "We are the first in Germany,"

he says. "There are some biotech companies off-campus near institutes. But what has never happened before is to set up both academics and a company under one roof."

Brüstle hopes that Life & Brain will incubate bright ideas from the university's academics. The company will lend technical and business support to these ideas and, where appropriate, hire academics to work full-time on their commercial application.



Oliver Brüstle: bringing a business edge to biology.

if you do not involve business."

Not everyone in Germany agrees with this approach. "We have proponents. We have opponents. That is how it is," says Matthias Winiger, president of the University of Bonn. But Brüstle dismisses some opposition as sour grapes over the amount of backing his company has obtained. "There was a lot of jealousy," he says. "A lot of money was raised."

Brüstle conceived Life & Brain with two other Bonn academics, including epilepsy researcher Christian Elger, now the company's scientific co-director with Brüstle. Each owns a 14% stake. A fourth scientist with start-up experience was recruited as co-founder, receiving a 13% stake. Bonn University and its medical school each took 20% stakes, with 5% reserved for other medical-school faculty.

Winiger, who became president of the university after Life & Brain got started, says he supports the decision to back it. "I'm proud the university took the risk." The company is an experiment that "needs time to grow," he says.

The state government of North Rhine-Westphalia is paying Life & Brain's rent for 15 years, at an estimated cost of €20 million. Lab equipment came from a €15-million grant from the compensation fund created to help Bonn recover from the 1999 transfer of the federal government to Berlin. And the medical school has

pledged up to €4 million a year to Life & Brain until it becomes self-supporting.

Early last year, Brüstle and his team occupied the new 10,000-square-metre laboratory building, which now houses 130 professors, postdoctoral researchers, students, technicians and support staff. Of these, 10 work full time, and 15 part time, for Life & Brain.

The company is divided into four divisions, each headed by a university professor: genomics, modelling transgenic disease, neurocognition, and stem-cell technology and reconstructive neurobiology — the last led by Brüstle. "The idea is to make it all profitable," he says. But for now, the company depends on the medical school's backing. Brüstle concedes



revenue is only trickling in, mainly from laboratory work under contract.

In the near term, he hopes to make money from automated systems for maintaining embryonic stem-cell cultures, developed with Hamilton, a robotics company based in Bonaduz, Switzerland. But the business plan calls for the main revenue stream to come from individual projects at Life & Brain. "These spin-offs will no longer have an academic component, they will be totally commercial," he says.

Brüstle concedes that therapeutic applications from his field of human embryonic stem-cell research are at least ten years away. But he says some applications will have commercial value much sooner: in particular, the use of stem cells as models for studying disease. "There is a huge demand for human-cell-based disease models in biomedicine and drug development," he says.

Heavy load

Johannes Dichgans, who retired last year as head of neurology at the University of Tübingen, says he supports more academic collaboration with business, but has reservations about how professors end up using their time. "A normal day for me was 12 to 14 hours and that is typical in this country," he says. "If I was also involved in a commercial entity, then I would have been divided so thin I could not have done anything right." Tübingen, he notes, uses a different model, with companies housed off-campus and professors kept off their payrolls.

Dichgans is also doubtful about the idea of a

Look west

Brüstle, 43, did postdoctoral work at the National Institutes of Health in Bethesda, Maryland, and he badly wants German academics to learn from their American colleagues' example. He thinks that German universities smother the creative impulses of young scientists, and says he wants to create a 'revolving door,' giving scientists the freedom to follow their ideas towards commercialization and back again to start on new ideas.

"There is not enough fire when it comes to translating academic research into applications," he says. "The life sciences cannot reach their full potential in a university setting



Revolving doors: Life & Brain's home could act as a gateway between academia and industry.

revolving door between academia and business, arguing that universities cannot usually hold positions open for professors wishing to spend a few years in commerce. "The way the German system is now, it will not work except for maybe a few very excellent people," he says.

But Markus Nöthen, head of genomics at the medical school, says his work as an unpaid research manager at Life & Brain has made him a better professor. "You think more about how scientific research can be exploited commercially," he says. "That is not the way a scientist is typically trained in Germany."

Brüstle insists that academics involved in Life & Brain are not short-changing the university. "People who work at Life & Brain have to put in additional time, and no one can do it all by working 40 hours a week."

But, says Winiger, it is already apparent that Life & Brain will take longer to become self-supporting than envisaged, meaning that the university will have to provide "quite a significant amount of money" to keep it afloat.

In that light, university and medical-school officials are debating whether to increase their 40% stake — effectively taking control of Life & Brain from Brüstle and his co-founders.

"Given the strong public support that Life & Brain currently receives, we have to be open to such an option," says Brüstle. New ownership or not, he holds to his vision: "I am convinced this thing is going in the right direction." ■

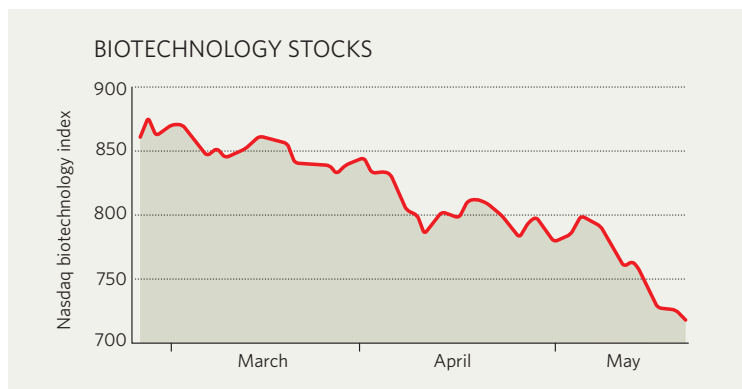
IN BRIEF

DYNAMIC DUO A new pairing of old drugs showed promising results in reducing a key marker of inflammation, a Boston biotechnology company has reported. CombinatoRx said that in a phase II trial, the steroid prednisolone combined with dipyridamole, an anti-clotting drug, significantly reduced levels of C-reactive protein in patients with chronic gum disease — although it did not reduce the size of gum lesions. The company is developing the drug combination, dubbed CRx-102, against a range of inflammatory diseases triggered by the immune system (see *Nature* 439, 390–391; 2006).

PIRATES' SLOW RETREAT The prevalence of pirated personal-computer software in China is slowly receding as the government there takes more steps to enforce intellectual property laws, says a global survey. The Business Software Alliance, an industry group representing software developers worldwide, reported that the percentage of computer software in China from pirated sources fell from 92% in 2003 to 90% in 2004 to 86% last year. It estimated that the value of the software pirated in China, if it had been paid for, would have been US\$3.9 billion. Estimated losses in the United States were \$6.9 billion, but only one fifth of software there was pirated.

OUT OF COTTON Syngenta, one of the world's main suppliers of genetically modified crop seed, is getting out of the cotton-seed business. On 23 May, the Swiss group announced that it will sell its interests in cotton to Delta and Pine Land, a seed company based in Scott, Mississippi, for an undisclosed sum. In April, Syngenta announced a collaborative agreement with DuPont subsidiary Pioneer Hi-Bred (see *Nature* 441, 149; 2006) and is expected to concentrate on corn and other food crops.

MARKET WATCH



This week Wood Mackenzie, an Edinburgh-based research and consulting firm, reviews recent trends in biotechnology stocks.

The Nasdaq biotechnology index has fallen steadily since the end of February, losing 13% of its value over the past eight weeks, and 8% since the start of the year. Early in the period charted above, general stock indices were rising, indicating that events specific to the biotech sector were behind the declines. However, all industries and market indices suffered in May, when fears over US inflation caused a steep drop across the board.

Paradoxically, the downward slide of the biotechnology index coincided with strong 2005 results by leading biotechnology companies. The past eight weeks have seen declines in the share values of most of them —

including big hitters such as Amgen, and Gilead whose stock fell by 7% and 11%, respectively, and Genzyme, which saw a 12% drop. Each company is heavily weighted in the index, so falls in their share value drag the whole index down.

The steady and uniform decline of the biotech index indicates a market correction, rather than a panic. After a 12-month rise, investors seem to have judged many biotechnology companies over-valued, and have cashed out to lock-in their gains.

The least affected of the major biotechnology stocks was Biogen Idec. The company's 2% drop over the past eight weeks follows an involuntary 'correction' of 40% in February 2005, when safety fears caused its candidate drug for multiple sclerosis, Tysabri, to be withdrawn.

Giving up the big questions when answers are in sight

SIR — Your Editorial and News Feature on the James Webb Space Telescope (JWST) correctly point out that large missions are difficult for NASA to accommodate during times of constrained budgets (*Nature* **440**, 127 and 140–143; 2006). But you do not explore what the community gives up if the largest missions such as JWST are sacrificed or delayed.

These ‘flagship’ missions are expensive because they are not merely large but scientifically transformational. One of the most intellectually significant, and humbling, scientific discoveries of our lifetime is that the atoms and forces of which we and our environment are made and that hold us together are but a trace contaminant of the true composition of the Universe, which is 96% dark matter and dark energy. These and similar discoveries simply could not have been made by a large number of low-cost space-science missions.

JWST has similar paradigm-breaking capabilities, in fields as diverse as the search for life-forming mechanisms on extra-solar planets and the birth of galaxies. In the era when it is scheduled to fly, all the current great observatories (Hubble, Chandra and Spitzer) will have completed their missions, or nearly so. A decision to give up on a small number of flagship missions is not merely a decision of economics and resource distribution, but a fundamental decision to avoid solving the grandest problems of science at a time when we know that they are within our reach.

Matt Mountain

Space Telescope Science Institute,
3700 San Martin Drive, Baltimore,
Maryland 21218, USA

Too many people are being displaced for Merowe dam

SIR — Your Special Report “Tide of censure for African dams” (*Nature* **440**, 393–394; 2006) considers the social consequences of the Merowe project in Sudan, which will displace up to 50,000 people. Can projects that involve such large population transfers possibly be justified?

The principal argument in favour of the Merowe project is a pressing need for electricity in Sudan. Merowe will provide electrical power corresponding to an average load of 625 MW (see en.wikipedia.org/wiki/Merowe_Dam).

With 50,000 displaced people for 625 MW of electricity, what I call the ‘displacement index’ for Merowe is 80 people per MW. By comparison, the smaller dam at Tignes

in the French Alps, whose construction in 1953 caused agitation and rebellion, displaced 384 people, for an average load of 103 MW. The corresponding ‘displacement index’ is 3.7 — one-twentieth that of Merowe.

This discrepancy is all the more sinister when one takes into account that more than 50 years have elapsed since Tignes. In the interval, the human suffering generated by resettlement schemes has, presumably, received increased recognition and respect.

In this time, friendlier power generating alternatives have also proven their merit. It is significant that a single present-day nuclear reactor (with an average load of 960 MW) would provide more power than the Merowe dam, with the displacement of very few people, if any. Would that not be the greener solution?

Maurice Guéron

15 rue Lakanal, 75015 Paris, France

Can the Internet save us from epidemics?

SIR — Kathleen Morrison, in News & Views (“Failure and how to avoid it” *Nature* **440**, 752–754; 2006), notes that societies have often prevented collapse by adopting new technological strategies. In today’s world, where one of the most-talked about prospects for collapse is an epidemic of infectious disease, it is worth remembering that perhaps we already have the technological strategy to avoid it — the Internet.

Remote working, made possible by the Internet (‘telepresence’), is already a key component of national and business pandemic plans. Telepresence can inhibit viral transmission by reducing human-to-human contact. Prepared organizations can leverage telepresence to allow continued productivity and functioning of supply chains during an outbreak.

Past societies often reacted to epidemics by bunching together, increasing density and transmission rates. In medieval Europe, for example, warring religious factions demonstrated solidarity in the face of a plague by marching together in the streets. And Native Americans expressed goodwill by gathering in the teepees of those infected with smallpox. But if we are well-prepared when an epidemic arrives, we can fluidly shift into a self-quarantined, telepresent society in which microbes fail by dint of host sparseness.

Whatever the social ills of increased isolation in our computer age, they may bode worse for the microbes than for us.

David M. Eagleman

Department of Neurobiology and
Anatomy, University of Texas, Houston,
Texas 77030, USA

Plant databases linked for botanists and gardeners

SIR — Your Editorial “The constant gardeners” (*Nature* **440**, 845; 2006) calls for closer cooperation between botanical gardens to integrate information about plants and make it accessible to all botanists. One such project, PlantCollections, is already under way. PlantCollections links the databases of 16 botanical gardens across the United States, using distributed queries on the Web. This approach allows organizations to keep their own database formats while sharing information. It uses software that translates an Internet query into the unique formats used by each of the 16 botanical gardens, collates the results and presents them to the user as a single document. The application is inexpensive to deploy and easy to maintain, so additional gardens can be added with minimal costs.

Information about living plant collections, herbarium specimens, long-term stored seeds, DNA samples and image holdings within each institution is shared in PlantCollections in a format that everyday gardeners as well as research scientists can understand. A list of participating botanical gardens and arboreta will be available at www.publicgardens.org. Further information on joining the project can be obtained from the author.

Boyce Tankersley

Chicago Botanic Garden, 1000 Lake Cook Road,
Glencoe, Illinois 60022, USA

People of the past should not be called primitive

SIR — Your Editorial “Rightful owners” (*Nature* **440**, 716; 2006) and Special Report “Guinea experts cry foul on tribal exhibits” (*Nature* **440**, 722–723; 2006) help bring to light the importance and challenges of repatriation and reclamation. But in them, you describe the makers of these antique South Pacific artefacts as “primitive”: a term that has been recognized as inappropriate by anthropologists for decades.

Referring to the communities that made this art as “primitive” implies that they were somehow beneath Western societies. Cultures developed uniquely across the world — these differences do not make one society better than another.

The word “primitive” reinforces colonial ideas of the innate rights of western societies to control others, and maintain their economic and cultural hold over these groups today.

Megan Marie McCullen

Department of Anthropology, Michigan State
University, 354 Baker Hall, East Lansing,
Michigan 48824, USA

BOOKS & ARTS

Sex and violence in rock art

Are cave paintings really little more than the testosterone-fuelled scribbles of young men?

The Nature of Paleolithic Art

by R. Dale Guthrie

University of Chicago Press: 2006. 520 pp. \$45

Paul G. Bahn

It is an odd fact that the art of the last Ice Age (the Upper Palaeolithic) is characterized by its numerous stylized or naturalistic animal images, and yet its study has rarely involved animal ethologists, apart from an occasional article by specialists in bison or big cats, or by a veterinarian keen to argue that some of the depicted animals were dead or dying. A palaeobiologist has now carried out a long, in-depth study of this art, and the result is *The Nature of Paleolithic Art*, a massive and unusual book whose witty title succinctly expresses its approach.

Dale Guthrie has in the past equated much palaeolithic art with the cheesecake of erotic magazines, and the trophies of hunting magazines. Here he presents these beliefs in far more depth and detail. He also sets the production of palaeolithic art into the context of the environment and social life, focusing particularly on life, love and death in the Ice Age. Much of this study is enlightening and valuable, drawing on the author's wealth of experience in cold environments and with hunting cultures.

Unfortunately, his reading of the images is extremely literal, excessively so in my view. In his quest for "clear expressions of the artists' hunting preoccupations", he thinks he can recognize depictions of "males with hunting weapons such as spears or spear-throwers", of hunting parties with men "being attacked by wounded game or dangerous carnivores", of nosebleeds and defaecations, and even of bleeding puncture wounds without weapons. People often see what they want to see in rock art, and I think it is safe to say that few of Guthrie's interpretations would be readily accepted by most specialists in Ice Age art.

The fundamental problem with these claims is that they are based, as so often in theories about rock art, on a small number of carefully selected examples, which are presented as typical. Some of his statements are surely exaggerations; in particular, declarations such as "images of rotund women, vulvae, and men with erections are found at many sites" or "giant penises ... occur throughout Paleolithic



Horse-play? Debate rages over the motivation for cave paintings of animals, such as this one at Lascaux.

art". Unfortunately, such imagery is far rarer and less characteristic than that presented here.

For example, if hunting and "testosterone themes" (to use Guthrie's term) were indeed one of the major causes of the art, surely one would find numerous hunting scenes. Yet in the tens of thousands of images, there is not a single clear depiction of a hunt, and the scenes highlighted in the relevant section of the book are sketchy and ambiguous, and lead Guthrie in some cases to provide his own interpretative drawings, dramatic 'reconstructions' of what he thinks is represented.

Regarding the highly enigmatic 'shaft scene' of Lascaux, he writes: "Surely there is little argument over what this scene is about," whereas in fact one would be hard pressed to find a depiction over which there has been more argument! Similarly, for Guthrie the Laussel 'playing card' figure "is clearly a couple copulating", but again it would be hard to find a vaguer and more enigmatic depiction. His interpretation of 'chimney' and 'Placard' signs as stylized women with legs spread is startling and novel, to say the least.

Another major emphasis of the book is the

author's conviction that works by youngsters constitute more of the art than had hitherto been proposed. This emphasis has led to regrettable and sensationalist headlines such as "Cave paintings are graffiti by prehistoric yobs", from the UK newspaper *The Independent on Sunday*, with much of the art now attributed to "sexually charged, intoxicated teenagers intent on vandalism". Guthrie's position is more sober, being based on new analyses indicating that many of the hand-stencils were done by young males. This is an interesting piece of evidence, but one wonders how far the results can safely be extrapolated to figurative Palaeolithic art. It has long been known that almost all the surviving footprints in the caves are those of children and adolescents, but that does not necessarily mean that they were also the artists, let alone that much of the art can be ascribed to sniggering youngsters.

One of the most welcome aspects of Guthrie's book is its determined shunning of the fantasies and speculations published in recent years about the role of so-called 'shaman-artists', 'trances' and 'altered states of consciousness' in the production of cave

art. Guthrie rightly states that the 'magico-religious' approach, as he calls it, has generated confusion and error, and has "resulted in a derailment of rock art research". It is worth noting that, with a lone exception, not one specialist in Ice Age art takes these shamanistic notions seriously at all. On the other hand, there can be no doubt that some kind of profound religious motivation — however one defines it — does lie behind some cave art. It cannot all be attributed to sex, hunting and teenage scribbling.

In short, this book attempts a far too literal

reading of much Palaeolithic art, and contains a great deal of wishful thinking in exaggerating the abundance, clarity and supposed ubiquity of hunting and sexual imagery in the art. Nevertheless, it provides a great number of interesting insights into the nature and behaviour of the species depicted, including humans, and is undeniably thought-provoking and challenging. I can recommend it highly, despite my reservations.

Paul G. Bahn is the author of *Journey Through the Ice Age* and *The Cambridge Illustrated History of Prehistoric Art*.

and Rayleigh. Ever sensitive to the cause of women's education, on the formation of the association governing Newnham College in 1880, Cayley became its first president. When he died at home in 1895 he had every reason to be content at a life well run.

But what of his mathematics? Many will be familiar with the beautiful Cayley–Hamilton theorem (a square matrix satisfies its own characteristic equation) and Cayley's theorem in group theory (any abstract group is isomorphic with a group of transformations). Students of phylogeny are grateful for his enumeration of trees, originally in connection with chemical formulae. But pure mathematics was his bent. His magnum opus was invariant theory, whose slow eclipse has taken Cayley with it. His *Collected Works* in 13 volumes containing 967 items (all single-author) awaits the diligent enquirer.

If one takes a polynomial like $ax^2 + 2bx + c$ and adds a value u , say, to x , resulting in a new polynomial $a'x^2 + 2b'x + c'$, it turns out that $a'c' - b'^2 = ac - b^2$. In other words, the function of the coefficients typified by $ac - b^2$ is an invariant. There are endless polynomials (Cayley called them 'quantics'), with different orders (quadric, cubic, quartic, and so on) and different numbers of variables (binary, ternary, quaternary, and so on). Boole found an invariant of the binary quintic but then decided that the theory was a "peculiar and rather isolated branch of analysis" and gave up.

But Cayley, and J. J. Sylvester and George Salmon in Dublin, pressed on. Although Crilly does his best to keep track of their developments, the task is nigh impossible at the level of a general biography. We are left to admire Cayley the mathematician at a distance, but Cayley the man we can admire more closely, a painstaking industrious beacon of the Victorian age. Thomson thought he ought to have been made 'mathematician laureate', while Maxwell serenaded him "whose soul, too large for vulgar space/In n dimensions flourished unrestricted".

Thus the real subject of Crilly's monumental biography is the surrounding galaxy of British mathematicians listed above and the milieu in which they operated. Centred on Cambridge, with outposts in London, Dublin and Edinburgh and good continental connections, this was the generation that took the torch of British mathematics from George Peacock and the other 'analyticals', and with it illuminated modern algebra and many other mathematical developments. Crilly has researched this period exhaustively, producing a work which will be the starting point for biographies yet to come. Might he himself now tackle Peacock, fascinating for his part in the reform of Cambridge as much as for his advocacy of modern algebra? We can only hope so.

A.W. F. Edwards is retired, but is still a fellow of Gonville and Caius College, University of Cambridge, Cambridge CB2 1TA, UK.

The forgotten mathematician

Arthur Cayley: Mathematician Laureate of the Victorian Age

by Tony Crilly

Johns Hopkins University Press: 2005.

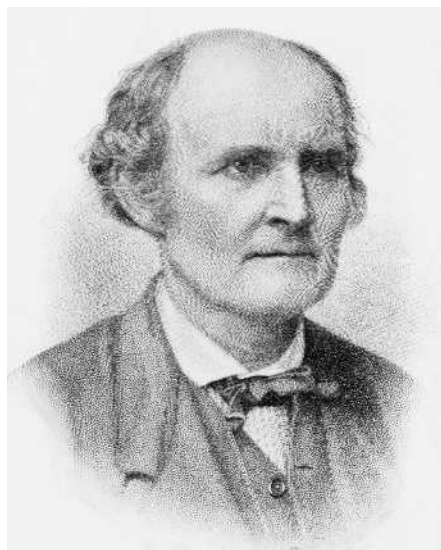
784 pp. \$69.95, £46.50

A. W. F. Edwards

It is not surprising that Arthur Cayley is a forgotten figure of the Victorian age, but it is wholly admirable that a modern biographer should enquire into his life and work so that we can judge him for ourselves. When Alexander Macfarlane gave his lectures on *Ten British Mathematicians* in 1901–04 he included Cayley; a century later, Ivor Grattan-Guinness edited *Landmark Writings in Western Mathematics 1640 to 1940* and Cayley was missing.

For those mathematicians and scientists not themselves destined to make fundamental contributions to their subjects, there is nothing more intellectually rewarding than the deep study of the work of those who have, what they achieved and how they came to achieve it. Biographers are needed who are proficient in the fields of enquiry that their subjects pursued, and Cayley is fortunate in having captured the attention of Tony Crilly. The problem then arises as to the level at which the subject's contribution is to be expounded: is Cayley to be revealed to mathematicians, or to a general scientific readership unversed in modern algebra? Crilly falls between the two stools. Mathematicians will regret the lack of mathematical detail, while others will risk losing themselves in a nebulous matrix of unfamiliar words (although mitigated by the author's exhaustive "Glossary of Mathematical Terms").

Many *Nature* readers will know about George Peacock, Cayley's teacher and a leading light in the Analytical Society, about William Thomson (Lord Kelvin), Lord Rayleigh and G. G. Stokes; we have read about James Clerk Maxwell and John Couch Adams, and enjoyed biographies of the two remarkable Georges, Boole and Green; we know of William Hamilton and his quaternions, and John Venn and his diagram. Augustus De Morgan, P. G. Tait,



Yesterday's man: over the past century, Arthur Cayley has been eclipsed by his colleagues.

H. J. S. Smith, J. W. L. Glaisher, R. L. Ellis and the indefatigable Isaac Todhunter are not unknown to us, and who has not heard of Charles Dodgson and of Francis Galton (whom Cayley taught)? All these men are listed in a helpful appendix, "Arthur Cayley's Social Circle". But who was Arthur Cayley?

Born in 1821 and senior wrangler in the Cambridge Mathematical Tripos in 1842, when he was elected a fellow of Trinity College, Cayley went to London to train as a barrister, being admitted to the bar in 1849. But he concentrated more on his mathematical than his legal work. Elected a fellow of the Royal Society in 1852, he was rewarded with the new Sadlerian professorship of pure mathematics at Cambridge in 1863, which he held for the rest of his life. He played major roles in the London Mathematical Society, the Royal Astronomical Society, and the British Association.

Cayley also led a full life in the affairs of the University of Cambridge and Trinity College. In 1888 he was granted an honorary ScD at the same congregation as Stokes, Adams

Red in the head

Seeing Red: A Study in Consciousness

by Nicholas Humphrey

Belknap Press: 2006. 160 pp.

\$19.95, £12.95, €18.50

Josh Weisberg

As a graduate student at the University of Cambridge in the late 1960s, psychologist Nicholas Humphrey discovered that a monkey named Helen, whose visual cortex had been surgically removed, could nonetheless interact visually with her environment. Helen's condition — later termed 'blindsight' by Humphrey's research supervisor Lawrence Weiskrantz when it was found in humans — became one of the most widely discussed bits of empirical data in the burgeoning field of consciousness studies. Blindsighted subjects guess at well above chance about the shape, orientation and even colour of stimuli that they insist they cannot see. The syndrome presents a kind of vision without conscious qualitative 'feel' — a provocative dissociation between the function of vision and its first-person appearance.

Since then, Humphrey has written extensively and insightfully about consciousness. His most recent work, a slim and elegant volume entitled *Seeing Red*, provides a charming, if brief, summary of his current views, blending themes culled from psychology, philosophy, and even art and poetry. It also offers intriguing speculations on the evolutionary function of consciousness. According to Humphrey, consciousness evolved to appear inexplicable, creating the impression that we are more than mere physical machines. Its mystery is its evolutionary *raison d'être*.

Humphrey holds that blindsight demonstrates the dissociability of sensation (the feel of experience) and perception (our awareness of the world). He adopts the bold suggestion that sensation is not part of the causal chain leading to perception. Instead, he argues, sensation makes up a separate, more primitive system that plays no direct role in our perception of the world. Sensation is a self-contained evaluative activity — in Humphrey's terms, someone seeing red engages in the activity of 'redding'. Humphrey models this activity on bodily expressions such as smiles: when I'm happy, I smile; when red light tickles my eyes, I respond by redding. However, sensory activity isn't connected to outward expression. Instead, it loops inwards and privately informs the person of the activation of the sense organs. This prompts affective responses, but not perceptions of the world.

In stressing the analogy with expressive activity, Humphrey hopes to mould conscious experience into a form amenable to identification in terms of brain processes. And it is plausible that certain features of expression are shared with sensation. Both smiles and



Red rag to a bull? Nicholas Humphrey boldly argues for a separation of sensation and perception.

IMAGES.COM/CORBIS

sensations of red are 'owned' by the subject — for example, only you can smile your smile. And both smiles and sensations occur in association with specific body organs. However, we are aware of our sensations with an immediacy that is lacking in the case of smiles. This is part of what it is for those sensations to be conscious. Failing to explain this immediacy leaves us in the dark about why a sensation is conscious. Even if sensation is identified with re-entrant feedback loops in the brain, as Humphrey suggests, we still need to know how these loops make us immediately aware of our sensations.

Although this crucial question is left unresolved, Humphrey's reflections on the evolutionary function of consciousness present an intriguing suggestion: our sense of the inexplicability of consciousness may expose its function. The apparent mystery of conscious-

ness prompts us to see ourselves as more than mere biological machines, and so to strive all the more to preserve our existence. However, a simpler explanation of this sense of mystery might be the lack of direct access to the underlying mechanisms of consciousness. But it is in the nature of such 'just so' speculations that they are extremely difficult to confirm.

In the final analysis, the strength of Humphrey's book lies in its skilful blending of ideas from varied sources to stimulate new ways of thinking about consciousness. In effectively doing so, while presenting a fascinating window on the thought of a distinguished consciousness researcher, *Seeing Red* is a wonderful success. ■

Josh Weisberg is in the Program in Philosophy and Cognitive Science, Graduate Center, City University of New York, 365 Fifth Avenue, New York, New York 10016-4309, USA.

Energy with meaning

Dictionary of Energy

edited by Cutler J. Cleveland & Christopher Morris

Elsevier: 2005. 512 pp. \$85, £50, €75

Ehsan Masood

Where to go for a definition of peak oil? What is an environmental Kuznets curve, and why was OPEC established? The expanding field of energy studies at last has its own encyclopaedic dictionary. Elsevier's *Dictionary of Energy* is all you could want in a book of this kind. Comprehensive and readable, it is a treat for fact-checkers and casual browsers alike.

The dictionary's coverage includes the science of energy, economics, sociology, policy

and energy in world history. The list of entries is peppered with more than 100 commissioned short essays on an array of topics and concepts. Many of these have been written by some of the leading names in the business.

The essay on the theory that oil production is close to peaking, for example, carries the name of the theory's main proponent, Colin Campbell. Robert Costanza, one of the leaders of the field of ecological economics, has provided a short essay on natural capital. The late Charles Keeling contributed an essay on the Mauna Loa curve, which describes the rate of change of atmospheric carbon dioxide over time; Keeling's measurements from the Mauna Loa observatory in Hawaii were one

of the first signals of climate change.

There are also short biographies of individuals who have made a lasting contribution, such as Vladimir Zworykin, the Russian-born physicist who patented the use of magnetic fields to guide cathode rays to produce fluorescent images on a screen.

The dictionary lacks a table of contents,

which means that the page after Zworykin's entry is both a surprise and a real delight: it is a collection of energy-related quotations (both funny and serious) from the past two centuries. Winston Churchill, Homer Simpson and Zaki Yamani, Saudi Arabia's former oil minister, all make an appearance here. So too does Golda Meir, Israel's formidable prime

minister in the early 1970s. "Let me tell you something we Israelis have against Moses," she is quoted as saying. "He took us 40 years through the desert in order to bring us to the one spot in the Middle East that has no oil." ■ Ehsan Masood is a freelance journalist and editor (with Daniel Schaffer) of *Dry: Life Without Water* (Harvard University Press, 2006).

Artists on a mission

An exhibition in London reflects on the dangers of climate change.

Colin Martin

The more the media bombard us with news of climate change, and the warmer the planet becomes, the greater is our nostalgia for the colder, whiter winters of the past. Peter Høeg's characterization of types of snowfall in his novel *Miss Smilla's Feeling for Snow* (Harvill, 1993), for example, captivated contemporary readers.

But the bigger picture — the social and global implications of accelerating climate change — has higher stakes. This is reflected in the output of the Cape Farewell project, in an exhibition called *The Ship: The Art of Climate Change*, which opens this Saturday (3 June) at the Natural History Museum in London. It runs until 3 September, after which some of the artworks can be seen at the Liverpool Biennial until November.

Over the past five years, Cape Farewell has organized three Arctic expeditions on the schooner *Noorderlicht*, bringing groups of artists, architects and writers, together with scientists and educators, to the Svalbard archipelago at 79° north. The objective is to raise public awareness of climate change by generating art installations and creative writing to complement and enliven the coldly objective data and dry scientific vocabulary of polar scientists and meteorologists.

"We intend to communicate through artworks our understanding of the changing climate on a human scale, so our individual lives can have meaning in what is a global problem," says David Buckland, artist and project director. Writer and participant Ian McEwan, a former Booker prize winner, has coined a memorable phrase, "the hot breath of our civilization", which pinpoints the problem as effectively as reams of scientific data and may stimulate more thought and action by individuals to try to halt climate change.

Some of the artists' responses to their Cape Farewell experiences were previously



Ice Towers by Peter Clegg and Antony Gormley could be seen in *The Ice Garden* in Oxford last December.

exhibited in *The Ice Garden*, an installation staged during four cold nights last December in the Clarendon quad next to the Bodleian Library in Oxford. Illuminated by a huge Moon, the setting evoked a seventeenth-century frost fair, like those held on the frozen Thames in London. Scrolling text by McEwan, projected onto an upper storey of the library, highlighted our man-made problem in lucid prose. Among other works were the sentinel-like *Ice Towers*, constructed by architect Peter Clegg and artist Antony Gormley. Each 0.54-cubic metre block of ice used in the piece represents the volume occupied by a kilogram of carbon dioxide at atmospheric pressure.

Soundscape, recorded by Max Eastley on land and under the sea at Spitsbergen, resonated eerily throughout the quad, providing an aural context for the visual works. Heather Ackroyd and Dan Harvey recreated their *Ice Lens* sculpture, concentrating artificial light through it to create a hot spot, recorded by thermal imaging. Their earlier attempt in the Arctic to scorch a sheet of paper by directing the

Sun's rays through a prototype had failed because the autumn rays were too weak.

Ice does not last long in urban summers, so in London Ackroyd and Harvey will exhibit a six-metre-long sculpture created from a Minke whale skeleton encrusted with alum crystals. Buckland will exhibit his photographs of text projected on to glacial ice, *Burning Ice*, *Sadness Melts* and *The Cold Library of Ice*. The Cape Farewell message remains the same, even though the artworks shown in London will not use ice as a physical medium.

The Cape Farewell project is a call for action, as well as a wake-up call. Coinciding with the London exhibition is a four-day conference over 11–14 July, Student Summit 2006: Climate Change, which aims to increase awareness and inspire advocacy among young adults, acknowledging that they represent our best chance of reducing carbon use. "On our side we have our rationality, which finds its highest expression and formalization in good science," says McEwan. "And we have a talent for working together — when it suits us."

Colin Martin is a writer based in London.

NEWS & VIEWS

CLIMATE CHANGE

The Arctic tells its story

Heather M. Stoll

The Arctic is one of the sensitive pressure points for Earth's climate. A new sediment core reveals much more about the region's role in a long-term transition from 'greenhouse' to 'icehouse' conditions.

Just as a white T-shirt keeps us cooler on a hot sunny day, a blanket of reflective snow and sea ice over the polar regions reduces the amount of sunlight the Earth absorbs. Ice and snow cover grows in response to cooling temperatures, and so amplifies climate change. But the Arctic and the Antarctic were not always frozen and barren. Over the past 55 million years or so, the Earth has experienced a major cooling, from a greenhouse climate to the current icehouse climate. The results of a remarkable scientific project in the Arctic¹, as reported in three papers in this issue^{2–4}, provide a detailed picture of the Arctic's role in the long-term cooling and its response during oscillations of the preceding greenhouse climates.

Despite the effect that the Arctic has on global climate — both through its icy reflection of the Sun's gaze and its efficient embrace of warm, salty currents from lower latitudes — its role in the climate transition has been a puzzle because no long geological record had been recovered from the central Arctic. Our understanding of Arctic climate evolution came from studies of cores and outcrops thousands of kilometres away. An astute strategy combining ice-breakers and drilling rigs (Fig. 1) has finally succeeded in getting the story from the horse's mouth. The Arctic coring expedition recovered a remarkable climate record spanning the past 55 million years from the central Arctic, and the inferences drawn from that record are described in the new papers^{2–4}.

Declining atmospheric CO₂ has long been envisaged as a culprit for the past 55 million years of cooling climates, and the consequent transition from an ice-free world to one with large ice sheets on Greenland and a frozen Arctic, as well as with a deep-frozen Antarctica. Indeed, reconstructions of past atmospheric CO₂ concentrations, based on isotopic markers from marine algae⁵, show a dramatic drop in atmospheric CO₂ between about 45 million and 25 million years ago that corresponds quite well to the onset of major global cooling (Fig. 2a,b, overleaf). However, although the onset of glaciation and sea ice in Antarctica about 43 million years ago matched the onset of global cooling and CO₂ decline, the Arctic



Figure 1 | Core contributions. Three ships, seen here on location, that worked together on the Arctic coring expedition: the ice-breaker/drillship *Vidar Viking*, and the ice-breakers *Oden* and *Sovetskiy Soyuz*.

seemed to march to a different drummer. Arctic cooling, with concomitant growth of ice sheets and sea ice, seemed to hold off for tens of millions of years, until about 2–3 million years ago when pebbles carried by icebergs first appeared in North Atlantic sediments.

To explain the long delay between Antarctic and Arctic glaciation, palaeoclimatologists suggested that Antarctic ice sheets started to grow not because of global climate factors such as CO₂, but in response to regional changes in the Southern Hemisphere. When the continents of South America and Australia drifted away from Antarctica, warm currents could no longer follow the coastline and make it all the way to Antarctica. Yet recent models suggested that cutting off currents alone would not have cooled Antarctica enough to start ice-sheet build-up⁶.

The puzzle is solved nicely with the new record reported by Moran *et al.*² (page 601), which shows that Arctic ice developed much earlier than previously believed. Pebbles

carried to the middle of the Arctic basin by icebergs appear by 45 million years ago, about the same time as around Antarctica (Fig. 2c). Abundant sand and iceberg-delivered pebbles confirm that sea ice and icebergs were widespread in the Arctic by 14 million years ago. This earlier onset of Arctic glaciation, synchronous with glaciation in Antarctica, supports the idea that changes in atmospheric CO₂ were the major driver in initiating glaciation in both hemispheres.

Although Moran and colleagues' paper² reveals a closer link between atmospheric CO₂ and the greenhouse–icehouse transition, results from older sediments in the Arctic core expose some surprising gaps in our understanding of the workings of climate in a greenhouse world. Sluijs *et al.*³ (page 610) report that 55 million years ago Arctic summertime surface-ocean temperatures were as high as 18 °C. Such temperatures are comparable to those of the modern summer ocean on the French coast at Brittany (where hardy

M. JAKOBSSON/ODP



50 YEARS AGO

The United Kingdom Atomic Energy Authority is holding a one-day symposium on controlled thermonuclear energy on June 4 at Harwell which will be on a classified basis. Invitations have been sent to representatives of industrial research laboratories in addition to government research establishments and British universities... The object of the symposium is to encourage fundamental research work along the lines which will assist the work of the Authority in this field. Research on the control of the thermonuclear reaction has been going on at Harwell for some time. The conference will be held in private, and no statement will be issued afterwards.

From *Nature* 2 June 1956.

100 YEARS AGO

The British Woodlice — At present in England there are only two dozen of these little land crustaceans on record. The number, combined with their love of obscurity, may remind us of the regal feast at which four-and-twenty blackbirds were served concealed in a pastry. When the pie was opened, the birds began to sing. In correspondence with the daintiness of such a dish, the apostles of oecology are now earnestly trying to persuade society that all nature is tuneful... The bright little volume under review is an excellent example of what can be done under the new impulse given to the old practice of "nature-study"... The many scurrilous colloquial terms that have been applied to these terrestrial isopods have, to the ordinary observer, obscured the fact that they are really made of one flesh and blood with the epicure's cherished treasures, the lobster and the prawn. Their use medicinally in old times would probably have been robbed of half its charm had this been understood, since in those days curative agencies seem to have been valued in proportion to the pain and disgust they inflicted on the patient... One may wonder whether the man who first ate a shrimp thought himself a hero!

From *Nature* 31 May 1906.

souls even go swimming). Most importantly, climate models for 55 million years ago don't come close to simulating such warm waters, even when reflective ice sheets are left out and atmospheric CO₂ levels are pumped up to 2,000 parts per million — nearly ten times the levels before the Industrial Revolution. Clearly, CO₂ is not the only driver of the extreme polar warmth. Something is missing from the models' climate simulation. The authors propose that that 'something' is another greenhouse agent, clouds of frozen water vapour in the lower stratosphere of polar regions. Like greenhouse gases in the atmosphere, these ice crystals trap part of the energy that Earth emits back to space, keeping Earth's surface warmer in polar regions.

Sluijs *et al.*³ also provide intriguing results for a dramatic burst of intense warming — the Palaeocene–Eocene Thermal Maximum — that occurred 55 million years ago. This 'Palaeocene supergreenhouse' is believed to have been caused by a massive release of carbon to the oceans and atmosphere, either from methane present in deep-sea sediments or as organic carbon vaporized by volcanism during the opening of the North Atlantic Ocean. In either case, the extra CO₂ in the atmosphere increased the greenhouse effect and warmed tropical temperatures by 4–5 °C (ref. 7). Sluijs *et al.* show that Arctic temperatures also soared, rising from 18 °C to 23 °C (Fig. 2c). Models are again unable to get the absolute temperatures right for the Arctic warming, but they do agree with the amount of Arctic warming caused by the CO₂ increase.

Unlike the situation observed over recent swings into and out of ice ages, where temperatures in the Arctic change by at least twice as much as those in the tropics, the warming in the Arctic during the Palaeocene supergreenhouse is about the same as that observed in tropical and subtropical regions. In this respect, paradoxically, this result confirms one aspect of our understanding of icehouse climates — that sea ice and ice sheets are responsible for the larger temperature swings in the polar regions, and in ice-free greenhouse climates the poles respond to climate changes just like everywhere else. This result reaffirms that, although the rate of CO₂ change and warming during the Palaeocene supergreenhouse may be similar to that expected in the coming centuries, in one respect future warming will be different — it will be strongly amplified at high latitudes by the reduction in snow and sea ice cover.

Finally, Brinkhuis *et al.*⁴ (page 606) provide a glimpse at the early operation of another crucial climate feedback in the Arctic, the relationship between heat transport and salinity. Today, warm salty currents feed into the Arctic. There, by releasing their heat, these waters become dense enough to sink into the deep ocean as North Atlantic Deepwater. Fluctuations in the intensity of these currents and deepwater formation participated in causing or amplifying the most abrupt climate

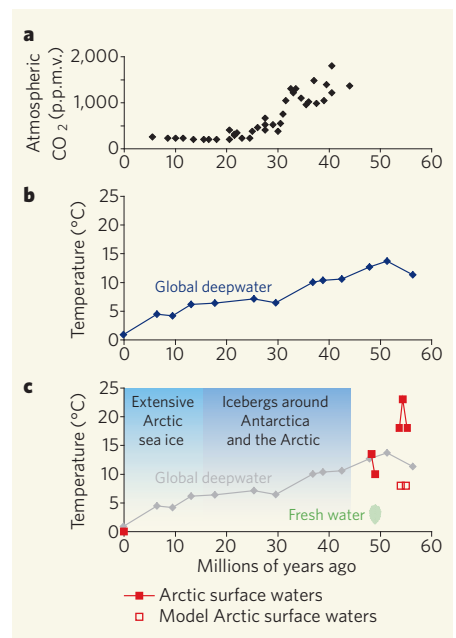


Figure 2 | Evolution of global and Arctic climate over the past 55 million years. **a, b,** Declining concentrations of atmospheric CO₂ (parts per million by volume; ref. 5) coincide with cooling inferred from deepwater temperatures⁸. **c,** Summary of the new results^{2–4}. The development of ice in the Arctic² occurred earlier than previously thought, and was synchronous with Antarctic glaciation. Fifty-five million years ago, during the Palaeocene–Eocene Thermal Maximum, summer temperatures of surface waters in the Arctic 'spiked' with a 5 °C rise. Temperatures before and during this event exceeded modelled temperatures by at least 10 °C (ref. 3). Finally, 49 million years ago, in a brief span of about 1 million years, the Arctic experienced an interval of unusually fresh surface waters⁴. This interval was truncated by influx of warm and salty currents originating in lower latitudes, which caused a rise in sea surface temperatures.

changes of the past tens of thousands of years.

Brinkhuis *et al.* identify an 800,000-year interval of time, 49 million years ago, when it seems that the Arctic may have been almost completely cut off from the inflow of warm and salty currents originating in lower latitudes. Without salty currents flowing in, the local excess of precipitation over evaporation created a freshwater environment (Fig. 2c) characterized by communities of aquatic ferns, *Azolla*, which today grow naturally only in waters with less than 0.2% salt. During this unique interval, pulses of freshwater even overflowed from the Arctic and carried remains of the ferns into the surrounding ocean basins.

The reign of the freshwater ferns in the Arctic ended abruptly 48.3 million years ago when waters became salty again. Significantly, Brinkhuis *et al.* show that the rise in salinity of the Arctic corresponded with a small but important rise in the temperature of Arctic waters — indicative of the entrance of warm and salty ocean currents from lower latitudes. As the younger sediments of the Arctic core release their secrets, we should find out

how the transport of warm ocean currents continued to evolve in the subsequent tens of millions of years, and how such changes relate to the falls in global temperatures over the past 20 million years.

Taken together, these papers^{2–4} resolve some enigmas about the climatic evolution of the Arctic. But we still need to sort out why the intensity of glaciation stepped up around 14 million years ago and increased further around 3 million years ago, because neither step was tied to decreases in atmospheric CO₂. Also, given that the modelled effect of CO₂ is not sufficient to explain the Palaeocene super-greenhouse and some other (possibly greenhouse) agent is required, then the demise of this other factor may also have affected the growth of ice sheets and the global cooling.

In all this, there's a particular challenge for those involved in climate modelling. If they can incorporate the processes causing the hot poles in the past, we will have even greater confidence in their predictions for the future. ■

Heather M. Stoll is in the Departamento de Geología, Universidad de Oviedo, Spain, and the Department of Geosciences, Williams College, Williamstown, Massachusetts 01267, USA.

e-mail: heather.m.stoll@williams.edu

1. www.ecord.org/exp/acex/302.html
2. Moran, K. *et al. Nature* **441**, 601–605 (2006).
3. Sluijs, A. *et al. Nature* **441**, 610–613 (2006).
4. Brinkhuis, H. *et al. Nature* **441**, 606–609 (2006).
5. Pagani, M. *et al. Science* **309**, 600–603 (2005).
6. Huber, M. & Nof, D. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 9–28 (2006).
7. Zachos, J. C. *et al. Science* **302**, 1151–1154 (2003).
8. Lear, C. H. *et al. Science* **287**, 269–272 (2000).

VIROLOGY

HIV goes nuclear

Min Li and Robert Craigie

HIV-1 replicates itself by integrating into its host cell's DNA. Studies in cell culture reveal that nuclear-membrane proteins aid engagement of the viral DNA with that of its host before integration.

Rather than going to the trouble of replicating its own genome, human immunodeficiency virus type 1 (HIV-1) inserts a DNA copy of its genome into that of its host so that the virus is reproduced during host-cell division (Fig. 1). The virus genome consists of RNA, so a DNA copy must first be made from the RNA template. The newly synthesized viral DNA forms part of a large complex in the cytoplasm of an infected cell, the pre-integration complex (PIC). This contains the viral integrase enzyme that splices the viral DNA into cellular DNA, as well as other viral and cellular proteins. To reach the cellular DNA, the PIC must cross the nuclear envelope that separates the cytoplasm and nucleoplasm. On page 641 of this issue, Jacque and Stevenson¹ show that HIV-1 PICs target a specific nuclear-membrane protein called emerlin, and that this is required for the viral DNA to engage with the host DNA before integration.

Unlike many viruses, HIV-1 and other lentiviruses can infect cells that are not dividing, so the viral DNA must cross the intact nuclear envelope to gain access to the cellular chromatin (the DNA and associated proteins that package it into chromosomes). This presents a formidable challenge, because the PIC is huge on a molecular scale — far larger than the diameter of nuclear pores. Thus, nuclear entry may be facilitated by first jettisoning some components of the PIC. Many protein components of the PIC carry signalling sequences that enable them to enter the nucleus, but the roles of these proteins and the

mechanism by which the PIC is imported into the nucleus remain controversial. Many cellular proteins have been reported to associate with PICs², but in most cases the functional consequences are unclear.

The nuclear envelope is composed of the inner and outer nuclear membranes, embedded pore complexes, and the lamina and associated proteins³. The nuclear lamina, which lies just inside the inner nuclear membrane, is a network of lamin proteins that, as part of the nuclear 'matrix', provides structural support for the nucleus. Proteins associated with the lamina include the lamin B receptor (LBR), lamin-associated polypeptides 1 and 2 (LAP1 and LAP2), emerlin and MAN1. LAP2, emerlin and MAN1 share a conserved structural domain (called the LEM domain) that binds to a protein known as the 'barrier-to-auto-integration' factor (BAF). BAF was first identified as a factor that blocks the self-destructive integration of murine leukaemia virus (MLV) DNA into its own genome *in vitro*⁴. This factor is a non-specific DNA-binding protein that can bridge strands of duplex DNA and may link chromatin to LEM-domain proteins at the nuclear envelope⁵. BAF is required for progression through the cell-division cycle⁶, and lack of BAF causes dividing cells to die⁷.

Jacque and Stevenson¹ examined the ability of HIV-1 to infect cells in which individual nuclear-envelope-associated proteins had been selectively depleted using a method called RNA interference (RNAi). This technique selectively targets messenger RNAs

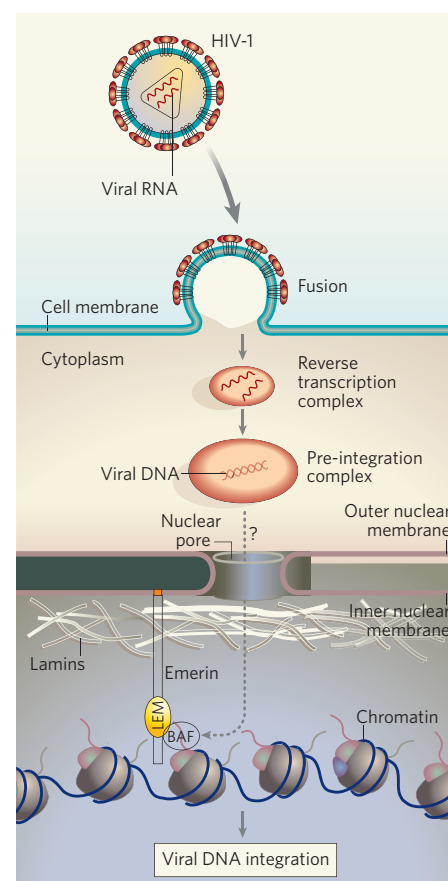


Figure 1 | HIV-1 integration into the host genome.

As HIV-1 is an RNA virus, to integrate into the genome of its host cell it must first be converted into DNA by reverse transcription. The viral DNA then forms part of the pre-integration complex with several proteins, including the integrase enzyme that splices it into the host DNA. Jacque and Stevenson¹ show that once the pre-integration complex has entered the nucleus (by an unknown mechanism), it targets a specific nuclear-membrane protein called emerlin. Emerlin forms part of the lamina beneath the nuclear membrane, where lamin proteins provide structural support for the nucleus. The host DNA may be associated with the lamina through the BAF protein, which binds to LEM-domain proteins such as emerlin. Jacque and Stevenson report that both emerlin and BAF are required for the viral DNA to integrate into the host DNA.

for degradation, stopping production of the encoded protein. The authors shrewdly studied the roles of the nuclear-envelope proteins in primary macrophages, one of the immune cell types infected by HIV. These cells do not divide, so this strategy bypassed the lethal effects of loss of BAF, and sidestepped potential cell-division roles of the other nuclear-envelope proteins tested.

Depletion of either emerlin or BAF severely impaired HIV-1 replication in infected macrophages. Notably, expression of these proteins using RNAi-resistant mRNAs restored replication, supporting the conclusion that the defect was the direct consequence of depleting each protein. The emerlin-depleted cells were fully susceptible to infection with a different

retrovirus (MLV), further indicating that the replication defect is specific to HIV-1 and not due to a widespread disruption of the normal cell physiology. Replication of MLV required BAF and LAP2 α , but not emerlin^{1,8}. So it would seem that, although both retroviruses recruit BAF, each virus also enlists different LEM-domain proteins.

What is the viral replication defect in cells depleted of either emerlin or BAF? In these cells, viral DNA was synthesized at normal levels but failed to integrate into the cellular genome. Biochemical subcellular fractionation experiments indicated that depletion of emerlin or BAF did not prevent the viral DNA from entering the nucleus, but that the viral DNA became associated with different nuclear fractions from normal. In control macrophages, most of the viral DNA was associated with the soluble chromatin fraction, whereas in cells depleted of BAF or emerlin it was mainly in the insoluble nuclear-matrix fraction. The identical infectivity defects caused by depletion of either emerlin or BAF suggest that these proteins have a cooperative role in HIV-1 infection, consistent with the known interaction between BAF and the LEM domain of emerlin.

The mechanism by which emerlin and BAF facilitate the proper nuclear localization of the HIV-1 PICs remains unknown. The association of emerlin with the PIC depends on BAF, which probably interacts with the viral DNA. But how does emerlin then influence the association of the HIV-1 PIC with the host chromatin? It may be that the work required to answer this question will also uncover other features of nuclear architecture. Although the organization of chromatin has been extensively studied at the level of the nucleosome (the smallest unit of DNA packaging), the global organization of chromatin within the nucleus is not well understood. However, the nucleus is clearly both highly compartmentalized and dynamic, and chromatin is intimately associated with the nuclear envelope. Perhaps we should not be surprised that there is more to accessing chromatin than simply crossing the nuclear envelope.

Min Li and Robert Craigie are in the Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: bobc@helix.nih.gov

DOWN'S SYNDROME

Critical genes in a critical region

Charles J. Epstein

The NFAT transcription factors activate the expression of many genes involved in the immune response and the development of a variety of tissues. They have now been implicated in Down's syndrome.

Down's syndrome is most commonly caused by the presence of an extra copy of the major portion of human chromosome 21. But how does the presence of an extra set of the roughly 200–300 genes on the chromosome give rise to the many abnormalities that characterize the condition? Because the pattern of abnormalities is so specific, one theory is that the 1.5-fold increase in the expression of some, if not all, of these genes is responsible¹.

In this issue, papers by Arron *et al.* (page 595)² and Gwack *et al.* (page 646)³ implicate two genes in the so-called Down's syndrome critical region (DSCR), a small segment of human chromosome 21, in causing the abnormalities found in Down's syndrome. Using diametrically opposed approaches, the groups reached this conclusion by a process that began with an interest in a family of four gene-regulatory factors called NFATc (for 'nuclear factor of activated T cells').

The regulation of various developmental pathways and of the immune response relies on processes that are activated by the entry of calcium into the cell, and the NFAT signalling

pathway mediates many of these processes. Following the influx of calcium, phosphate groups are removed from NFATc factors in the cytoplasm by the enzyme calcineurin. This allows NFATc to enter the nucleus and activate its target genes. However, once in the nucleus, NFATc can have phosphate groups added back to it by a kinase enzyme (phosphorylation), forcing it to return to the cytoplasm and halting its effects on the genes (Fig. 1a).

Arron *et al.*² came upon the possibility of a connection between the NFAT system and Down's syndrome by the serendipitous observation that mice lacking NFATc2 and NFATc4 have abnormalities of the skull and jawbone. These deformities are similar to those observed in Down's syndrome and in two mouse models of Down's syndrome (called Ts65Dn and Ts1Cje) that have an extra copy of part of the mouse chromosome most similar to human chromosome 21 (that is, they are trisomic)⁴. In addition, these and other mice lacking various NFATc family members, either singly or in combination, display abnormalities that are highly reminiscent of Down's

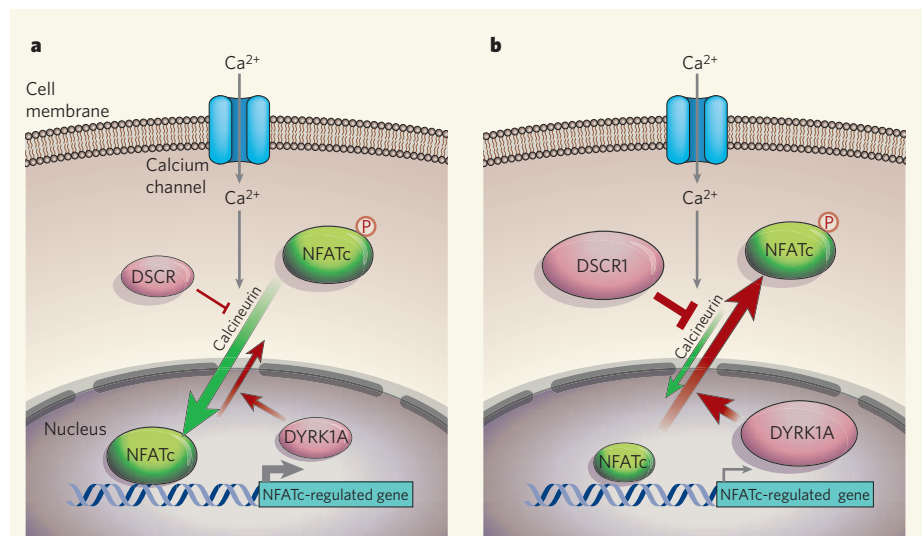


Figure 1 | NFAT signalling and Down's syndrome. Calcium signalling through the NFATc pathway mediates many developmental processes and the immune response. **a**, The entry of calcium ions into the cell activates the enzyme calcineurin to remove phosphate groups (P) from NFATc factors in the cytoplasm, allowing NFATc to enter the nucleus and activate its target genes. However, once in the nucleus, the NFATc can be phosphorylated, and so returns to the cytoplasm. Arron *et al.*² and Gwack *et al.*³ implicate the DSCR1 and DYRK1A proteins in regulating the levels of NFATc phosphorylation. **b**, The genes encoding DSCR1 and DYRK1A are found in the 'Down's syndrome critical region' of human chromosome 21, which has an extra copy in people with Down's syndrome. The increased expression of DSCR1 and DYRK1A disturbs the balance of NFATc phosphorylation, so that most of the protein is found in the cytoplasm^{2,3}. Thus, NFATc-dependent genes will not be properly regulated, which could markedly affect development. (Modified from Arron *et al.*², Supplementary Figure 1.)

- Jacque, J.-M. & Stevenson, M. *Nature* **441**, 641–645 (2006).
- Turlure, F., Devroe, E., Silver, P. A. & Engelman, A. *Front. Biosci.* **9**, 3187–3208 (2004).
- Gruenbaum, Y., Margalit, A., Goldman, R. D., Shumaker, D. K. & Wilson, K. L. *Nature Rev. Mol. Cell Biol.* **6**, 21–31 (2005).
- Lee, M. S. & Craigie, R. *Proc. Natl Acad. Sci. USA* **95**, 1528–1533 (1998).
- Segura-Totten, M. & Wilson, K. L. *Trends Cell Biol.* **14**, 261–266 (2004).
- Furukawa, K. *et al. J. Cell Sci.* **116**, 3811–3823 (2003).
- Margalit, A., Segura-Totten, M., Gruenbaum, Y. & Wilson, K. L. *Proc. Natl Acad. Sci. USA* **102**, 3290–3295 (2005).
- Suzuki, Y., Yang, H. & Craigie, R. *EMBO J.* **23**, 4670–4678 (2004).

syndrome: behavioural changes, decreased muscle strength, heart defects, annular pancreas, aganglionic megacolon, immune defects and placental vascular abnormalities.

So Arron *et al.*² looked for genes in the DSCR that could regulate NFATc. They identified *DYRK1A*, which encodes a kinase, and *DSCR1*, which encodes a known inhibitor of calcineurin. The *DYRK1A* and *DSCR1* proteins acted synergistically to block NFATc-regulated gene expression. The authors found that *DYRK1A* acts as a priming kinase that enables additional phosphorylation of NFATc by another kinase, glycogen synthase kinase 3, leading to NFATc inactivation. Furthermore, mice genetically engineered to have increased levels of *DYRK1A* alone, or of both *DYRK1A* and *DSCR1*, had abnormal heart-valve development. As expected, in these mice NFATc was mostly phosphorylated and found in the cytoplasm (Fig. 1b).

Gwack *et al.*³ reached a similar conclusion on the role of *DYRK1A* in regulating NFAT signalling by using an innovative approach in fruitfly cells to identify regulators of the mammalian NFAT pathway. Although these cells do not themselves possess NFATc factors, the pathway regulating their movement is present. This provided the basis for a genome-wide screen using interfering RNAs to inhibit the expression of specific genes. The authors identified *DYRK1A* and its relative *DYRK2* as direct regulators of NFATc phosphorylation, and, like Arron *et al.*, they showed that *DYRK1A* has a priming function. Gwack *et al.* note that *DYRK1A* and *DSCR1* occur in the DSCR, and suggest that their findings might aid the understanding of the immunological and neurological defects in Down's syndrome.

But does perturbation of the NFAT system explain the many developmental abnormalities of Down's syndrome? The difficulty in establishing this, as Arron *et al.* point out², is that the various genetically engineered mice show different developmental defects. The trisomic Ts65Dn and Ts1Cje mice, which express extra copies of *DYRK1A* and *DSCR1* (ref. 5), show few of the abnormalities found in mice deficient in the various NFATc factors. But the trisomic mice also show few of the abnormalities of Down's syndrome, although they do have cellular abnormalities that may be highly relevant in studying the condition^{6,7}. Such discrepancies may result from subtle variations in the expression of components of the NFATc regulatory system that occur because of the ways the relevant genes have been introduced or knocked out. Similarly, the discrepancies between Down's syndrome in humans and the mouse models may derive from species differences in gene expression and regulation.

Whatever their causes, these discrepancies make it more difficult to analyse directly the components of Down's syndrome that are not replicated in the trisomic mouse models. However, one feature that occurs in the human syndrome, NFATc-deficient mice and

trisomic mice is the abnormalities of the skull and jawbone that provided the original impetus for the search by Arron *et al.*^{2,4}. So a straightforward way to test whether increased *DYRK1A* and/or *DSCR1* expression is the cause of these defects would be to reduce the number of copies of either or both genes from three to two by mating trisomic mice with mice lacking the genes. A similar approach has already been used to explore the role of the APP protein in the brains of trisomic mice⁸.

Increased *DYRK1A* activity was already suspected to be a factor underlying the cognitive and muscular deficits in Down's syndrome^{9,10}. The work of Arron *et al.*² and Gwack *et al.*³ suggests another way that increased activity of *DYRK1A* and *DSCR1* may contribute not only to mental retardation, but also to many other features of Down's syndrome. ■

Charles J. Epstein is in the Department of Pediatrics and the Center for Human Genetics, University of California, San Francisco, Rock Hall 584B, 1550 4th Street, San Francisco, California 94143-2911, USA. e-mail: charles.epstein@ucsf.edu

1. Epstein, C. J. *The Consequences of Chromosome Imbalance: Principles, Mechanisms, and Models* (Cambridge Univ. Press, 1986).
2. Arron, J. R. *et al.* *Nature* **441**, 595–600 (2006).
3. Gwack, Y. *et al.* *Nature* **441**, 646–650 (2006).
4. Richtsmeier, J. T., Baxter, L. L. & Reeves, R. H. *Dev. Dyn.* **217**, 137–145 (2000).
5. Amano, K. *et al.* *Hum. Mol. Genet.* **212**, 1333–1340 (2004).
6. Belichenko, P. V. *et al.* *J. Comp. Neurol.* **480**, 281–298 (2004).
7. Kleschevnikov, A. M. *et al.* *J. Neurosci.* **24**, 8153–8160 (2004).
8. Cataldo, A. M. *et al.* *J. Neurosci.* **23**, 6788–6792 (2003).
9. Branchi, I. *et al.* *J. Neuropathol. Exp. Neurol.* **63**, 429–440 (2004).
10. Altafaj, X. *et al.* *Hum. Mol. Genet.* **10**, 1915–1923 (2001).

CLIMATE CHANGE

All in the game

Thomas Pfeiffer and Martin A. Nowak

It is in the public interest to keep Earth's climate on an even keel — the public, in this case, being all the world's population. Are you prepared to stake your own reputation on helping to improve matters?

Earlier this year, the German newspaper *Hamburger Abendblatt* published an advertisement¹ drawing attention to the impact of human activities on global climate:

Human activities have already demonstrably changed global climate, and further, much greater changes must be expected throughout this century. The emissions of CO₂ and other greenhouse gases will further accelerate global warming... Some future climatic consequences of human-induced CO₂ emissions, for example some warming and sea-level rise, cannot be prevented, and human societies will have to adapt to these changes. Other consequences can perhaps be prevented by reducing CO₂ emissions. Everyday measures can contribute to climate protection.

Nothing odd about that, you might think. But there was. Not only was the advert written by the director of the Max Planck Institute for Meteorology in Hamburg, Jochem Marotzke, but it was paid for by the proceeds of a 'climate game' that was designed to explore the willingness of individuals to invest in protecting Earth's climate (Fig. 1, overleaf). A paper² describing the climate game by Manfred Milinski, along with Marotzke and two other colleagues, was published simultaneously with the advert. What's the story behind that paper and the advert?

The Earth's climate is a 'public good'. It is shared by the entire human population, so it apparently does not pay for a single individual

to invest in protecting it: the direct benefit that an individual gains from his or her investment is much smaller than the costs. The Earth's climate is therefore vulnerable to overexploitation — it faces a 'tragedy of the commons'³. Experiments with public goods games are one way of exploring how such a tragedy might be avoided.

In a typical public goods game^{4–7}, members of a group can simultaneously invest money in a group project. The group's investment is increased by the experimenter, and the return is shared equally among all members of the group. In such a setting, it typically does not pay for a player to invest, because the player's direct return will tend to be a fraction of his or her investment. Assume, for example, that the group consists of three players, and the experimenter doubles the group's investment. Then, when one player invests (say) \$1, and the others invest nothing, that player gets back only two-thirds from his or her own investment, because that investment is doubled by the experimenter but then divided among all three players. Making an investment is an altruistic act in favour of the other players. The best strategy is to invest nothing in the group project but to benefit from the others' investment. It is therefore not surprising that, if played repeatedly, the amount of money invested in the group project declines rapidly³. But if all the members of a group instead continue to invest, the entire group is better off.

Max-Planck-Institut für Meteorologie

Max Planck Institute for Meteorology



Professor Jochem Marotzke, Geschäftsführender Direktor des Max-Planck-Instituts für Meteorologie in Hamburg:

„Der Mensch hat das globale Klima bereits nachweislich geändert, und weitere, wesentlich größere Änderungen sind für dieses Jahrhundert zu erwarten. Der Ausstoß von CO₂ und anderen Treibhausgasen wird die globale Erwärmung weiter verstärken. Als Folge müssen wir mit häufigerem Auftreten von Klima- und Wetterextremen über 15.000 Todesfällen allein in Frankreich, werden häufiger...

Verantwortlich für den Text: Professor Jochem Marotzke, Max-Planck-Institut für Meteorologie, Hamburg. Diese Anzeige wurde durch Spenden finanziert. Studenten der Universität Hamburg nahmen an einem „Klimaspiel“ teil, das die Bereitschaft der Teilnehmer erforschte, eigenes Geld für den Klimaschutz auszugeben. Für weitere Informationen siehe Pressemitteilungen „Gewinn muss nicht klimpern“ auf unserer Homepage www.mpimet.mpg.de sowie unter www.mpg.de auf der Homepage der Max-Planck-Gesellschaft.

Figure 1 | That ad — and how it was paid for.

Such cooperation between group members can be achieved by changing the rules of the game. For example, if players can punish or reward other players based on their behaviour in the public goods game, a high level of investment is maintained^{4,5,7}. Players who are known to make investments in the public good gain a good reputation; and that reputation translates into direct benefits of either receiving rewards or avoiding punishment.

Maintaining cooperation in these games seems far removed from addressing such problems as global warming. But Milinski and colleagues' climate game² is a step in that direction. As in previous studies, rounds of a public goods game were alternated with rounds of an 'indirect reciprocity game'^{8–11} in which players could reward other players based on their reputation. But in contrast to previous versions of the game, the yield from the public goods game was not given back to the group. Instead, it was paid into a 'climate fund' that was used to pay for the advert. Amazingly, the players still maintained a high level of cooperation in the public goods game, and in the indirect reciprocity game they preferentially rewarded individuals who invested in the climate fund.

The experiments revealed that reputation is essential in maintaining a high level of investment. In rounds where the investment was anonymous, the degree of cooperation was much smaller than in rounds where the investment was made public. Apparently, individuals invested in the climate game to maintain a good reputation and thereby increase the chances of being rewarded by the other players in the indirect reciprocity game. The main message of this study, then, is that whenever individual behaviour is relevant to the public good, it should be made public. Only then can an individual's regard for his or her own reputation be fully exploited.

How might these principles work in practice? Energy costs of individual households could, for example, be published by local authorities. Companies could be ranked according to their emissions and their investments in climate protection. The large costs that may arise from global climate change have to be communicated, even if the details are not yet fully understood. And even small investments in climate protection may help, such as

a slight reduction of room temperatures in winter or making the effort to use public rather than private transport, where available.

Let no one underestimate the difficulty in persuading individuals around the world, let

alone businesses and entire nations, about the necessity of putting the brakes on the human impact on climate. And cynics may sneer at the prospect of applying the principles of an academic exercise on a world stage. But like it or not, we are all involved in the very real game of climate change. It is one of the few games we cannot afford to lose.

Thomas Pfeiffer and Martin A. Nowak are in the Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, and Department of Mathematics, Harvard University, Cambridge, Massachusetts 02138, USA. e-mail: pfeiffer@fas.harvard.edu

1. *Hamburger Abendblatt* 20 (9 March 2006).
2. Milinski, M., Semmann, D., Krambeck, H.-J. & Marotzke, J. *Proc. Natl Acad. Sci. USA* **103**, 3994–3998 (2006).
3. Hardin, G. *Science* **162**, 1243–1248 (1968).
4. Fehr, E. & Gächter, S. *Nature* **415**, 137–140 (2002).
5. Milinski, M. et al. *Nature* **415**, 424–426 (2002).
6. Brandt, H., Hauert, C. & Sigmund, K. *Proc. Natl Acad. Sci. USA* **103**, 495–497 (2006).
7. Panchanatan, K. & Boyd, R. *Nature* **432**, 499–502 (2004).
8. Nowak, M. A. & Sigmund, K. *Nature* **393**, 573–577 (1998).
9. Wedekind, C. & Milinski, M. *Science* **288**, 850–852 (2000).
10. Nowak, M. A. & Sigmund, K. *Nature* **437**, 1291–1298 (2005).
11. Ohtsuki, H. & Iwasa, Y. *J. Theor. Biol.* **239**, 435–444 (2006).

PHYSICAL CHEMISTRY

Gas in a straitjacket

Susumu Kitagawa

Given a holding material with sufficiently small and uniform pores, gaseous oxygen can be made to form regular one-dimensional chains. That gives unprecedented insight into the properties of confined gases.

Porous materials are invaluable 'hosts' — that is, they are useful for storing, separating and investigating various guest compounds. Typical hosts are inorganic materials called zeolites (which contain a lattice structure) or so-called granular activated carbon, which possesses a large internal surface area to accommodate guests. Just recently, a third group of materials, metal-organic frameworks known as coordination polymers, have joined this party of excellent hosts^{1,2}. Writing in *Angewandte Chemie International Edition*, Takamizawa and colleagues³ report a canny use of a coordination polymer to store molecular oxygen (strictly, dioxygen, O₂) as an ordered array of three molecules (a trimer) or as a regular one-dimensional chain over a wide range of temperatures.

Molecular oxygen is a particularly welcome guest molecule, for several reasons. First, its so-called 'frontier orbitals' include unpaired electrons, and give rise to a rich array of reduction and oxidation reactions (involving the gain and loss of electrons, respectively). These electrons, through their unpaired spins, also make O₂ the smallest stable 'paramagnetic' molecule, meaning that it orients itself in a particular way in response to a magnetic field. Confining oxygen in an array would allow

currently unknown details of the interplay of localized atomic spins and delocalized electrons to be investigated in charge-transfer and redox reactions between the oxygen and the pore surfaces.

Second, the properties of oxygen in its ordered, solid form are particularly interesting: in various three-dimensional solid phases under different conditions, it both exhibits a form of magnetism known as antiferromagnetism (in which the spins of neighbouring electrons align pointing in opposite directions) and displays metallic conductivity and superconductivity⁴. In lower-dimensional, confined forms of oxygen, novel magnetic and optical properties are expected⁵. The problem is that oxygen is a gas under ambient conditions, and solidifies only at −218.9 °C. Investigation of ordered forms of oxygen at ambient temperatures therefore requires extremely high pressures.

Coordination polymers offer just the right tool to bring about an ordered state without applying energy in the form of pressure. The key to this capability is extremely small, regular 'ultramicropores' of less than 0.7 nanometres in diameter that allow guest molecules to be lined up uniformly and intimately within the polymer. Conventional hosts such as

PLANETARY SCIENCE

Polar roller

In July 2005, the Cassini space probe observed immense plumes of water vapour (pictured) spouting from an anomalously warm region near the southern pole of the saturnian moon Enceladus. That spectacular observation confirmed a long-held belief that Enceladus is the source of the icy grains that make up its home — the outer lane, or 'E ring', of Saturn's ring system.

But why are these plumes where they are? In this issue, Francis Nimmo and Robert T. Pappalardo

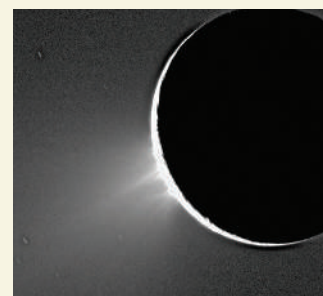
(*Nature* **441**, 614–616; 2006) show that the spout's polar location could, theoretically, be explained by the formation of a diapir, an upwelling region of low-density material, in Enceladus' interior. Simple gravitational considerations require that the moon swings round so that this less weighty, tectonically active region is aligned with the moon's axis of rotation — and so ends up at the pole.

There are caveats. The authors show that if the diapir is within

Enceladus' subsurface ice shell, then the moon's outermost layer, or lithosphere, must be relatively rigid. Otherwise, the upwelling material would simply bulge out from the surface. That region would then tend to migrate not towards the poles, but towards the moon's equator.

A diapir within Enceladus' silicate core, on the other hand, would cause a poleward reorientation. But it would preclude the existence of a global subsurface ocean: such an ocean would decouple the core gravitationally from the ice shell, preventing reorientation.

Nimmo and Pappalardo's low-



density diapir should be detectable through its effect on the moon's gravity field. Further analysis of data returned by radio instruments aboard Cassini should provide a conclusive test for its existence.

Paul Hanlon

NASA/JPL/SPACE SCI. INST.

activated carbons do not permit such structurally well-characterized arrangements: their larger pore sizes mean that the guest molecule always has some freedom to move, resulting in highly disordered arrays that must be modelled statistically.

Compared with conventional hosts, coordination polymers also possess unusual flexibility and their physical and chemical properties can be tuned by adding specific functional groups at the surface. Their molecules and ions can also be designed in minute detail. Complex frameworks can thus be built up, in which clusters, wires, ladders and other exotic molecular conformations can be contained. Molecules confined to these pore networks experience an unusually high, internally generated pressure⁶, giving rise to condensed assemblies whose properties differ from those of the bulk solid or those of assembled forms in an ambient environment.

Takamizawa *et al.*³ investigated oxygen trimers included along a channel constructed from a linear chain of a coordination polymer over a wide range of temperatures, from room temperature down to 10 K. They observed an evolution from a 'softer', more flexible assembly of oxygen trimers at higher temperatures to a rigid, three-dimensional ordered assembly at the lowest temperatures. Analysis of the geometry of the trimers showed that the internal distance between the trimer elements increased steadily with temperature, but that the distance between trimers first decreased to a minimum at around 90 K, and then increased once again. This indicated the presence of an intermediate, unstable phase at minimum trimer separation that could — according to measurements of magnetic correlations between the trimers involved — be treated as a quasi-one-dimensional chain of oxygen molecules.

Interestingly, the authors also show that this intermediate trimer structure changes its form under the influence of a magnetic field. Porous crystalline coordination polymers are often much more dynamic than is generally believed, and numerous studies have shown

that they possess a structural flexibility that can lead to physical changes when an external stimulus — in most cases the chemical stimulus of a guest molecule — is applied⁷. In Takamizawa and colleagues' case³, it seems that the host framework itself flips back and forth between states, allowing the oxygen trimer to change its form in response to the applied magnetic field. This demonstration follows the magnetically induced rearrangement of an included oxygen dimer last year⁸.

These results could lead to new methods for the remote control of molecular structures. Whereas molecular devices that respond to light have been readily synthesized, similar systems that respond to magnetic fields have not, as the energy contained in magnetic interactions is generally too small to drive structural transformations. Magnetically induced phase transitions in an assembly of included guest paramagnetic molecules offer a way round this

obstacle, as even a small stimulus could give rise to larger-scale structural change in both guest and host. In particular, guest molecular clusters could provide new quantum spin states that are strongly coupled to those of the polymer structure surrounding them.

Because Takamizawa and colleagues' polymer framework is so flexible, it can easily accommodate guest compounds that are not just paramagnetic, and thus responsive to magnetic fields, but also dipolar, and therefore sensitive to electric fields. These molecules range from heterodiatom molecules such as carbon monoxide (which is dipolar) and nitrogen oxide (paramagnetic and dipolar) to larger polyatomic molecules such as sulphur dioxide (dipolar) (Fig. 1). With so many guests to choose from, it is inevitable that new topics of discussion will arise concerning ultramicrospores⁹. What is emerging is a new science of 'coordination space' that studies the role of nanospaces in determining the chemical and physical functions of guest and host frameworks.

Small gas molecules such as O₂, CO, NO and SO₂ have, in their separate ways, important roles in gas separation and storage, catalysis and pollution control. An in-depth understanding of field-responsive properties of the hybridized gas assemblies offered by Takamizawa and colleagues could pave the way to new applications in environmental, energy and life sciences.

Susumu Kitagawa is in the Department of Synthetic Chemistry and Biological Chemistry, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510, Japan.
e-mail: kitagawa@sbchem.kyoto-u.ac.jp

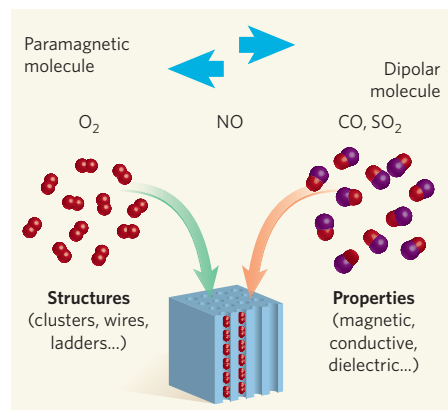


Figure 1 | Versatility of Takamizawa and colleagues' porous coordination polymer³.

Coordination polymers provide nanometre-sized vessels to accommodate paramagnetic and/or dipolar molecules in low-dimensional arrays. Their flexible framework means that the host polymers can change their form to better suit their guests, allowing molecular configurations of gas molecules to be created — and subsequently readily released — without the need for extreme temperatures or pressures.

1. Yaghi, O. M. *Nature* **423**, 705–713 (2003).
2. Ferey, G. *et al.* *Acc. Chem. Res.* **38**, 217–225 (2005).
3. Takamizawa, S. *et al.* *Angew. Chem. Int. Edn* **45**, 2216–2221 (2006).
4. Shimizu, K. *et al.* *Nature* **393**, 767–769 (1998).
5. Ritter, S. K. *Chem. Eng. News* **82** (50), 29–32 (2004).
6. Kitaura, R. *et al.* *Science* **298**, 2358–2361 (2002).
7. Bradshaw, D. *et al.* *Acc. Chem. Res.* **38**, 273–282 (2005).
8. Kobayashi, T. C. *et al.* *Prog. Theor. Phys. Suppl.* **159**, 271–279 (2005).
9. Langley, P. J. & Hulliger, J. *Chem. Soc. Rev.* **28**, 279–291 (1999).

BRIEF COMMUNICATIONS

Subsidence and flooding in New Orleans

A subsidence map of the city offers insight into the failure of the levees during Hurricane Katrina.

It has long been recognized that New Orleans is subsiding and is therefore susceptible to catastrophic flooding. Here we present a new subsidence map for the city, generated from space-based synthetic-aperture radar measurements, which reveals that parts of New Orleans underwent rapid subsidence in the three years before Hurricane Katrina struck in August 2005. One such area is next to the Mississippi River–Gulf Outlet (MRGO) canal, where levees failed during the peak storm surge: the map indicates that this weakness could be explained by subsidence of a metre or more since their construction.

To make the subsidence map, we used 33 scenes recorded from Canada's RADARSAT satellite. The technique involves interferometric phase comparison of 33 synthetic-aperture radar images taken at different times along the same nominal orbit, and exploits points on the ground that strongly reflect radar, termed 'permanent scatterers' (for details of the analysis, see supplementary information).

Figure 1 shows the average distance (range) changes of the permanent scatterers along the direction of radar illumination over the three years (2002–05) during which data were collected. It indicates that subsidence was

widespread during this period. All results presented here are range changes in the direction of radar illumination, but it is reasonable to assume that the principal motion was vertical. The mean and standard deviation in rate of range change for all the point targets was $-5.6 \pm 2.5 \text{ mm yr}^{-1}$ (negative rates indicate subsidence). The maximum rate observed was -29 mm yr^{-1} . If surface motion was purely vertical, the corresponding mean and maximum subsidence rates were 6.4 and 33 mm yr^{-1} , respectively. As global sea level is rising at an average rate of about 2 mm yr^{-1} (ref. 1), most of New Orleans is subsiding relative to the global mean sea level at an average rate of about 8 mm yr^{-1} .

Our data have exceptional spatial resolution, allowing assessment of the motion of individual buildings and other structures. Comparison with high-resolution aerial photographs shows that many of the scatterers are located at the intersection of a road and the wall of a building, where the road is roughly parallel to the satellite flight path (perpendicular to the radar look direction), or on building roofs (see supplementary information). The mean elevation of scatterers is therefore considerably higher than the mean ground

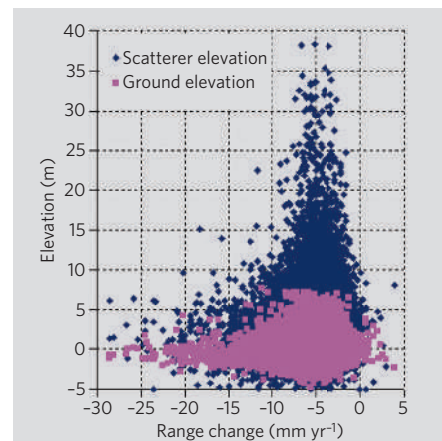


Figure 2 | Relation between range change of permanent scatterers and elevation. Negative range change indicates subsidence. The elevation of many permanent scatterers is several metres to several tens of metres above the land surface because many permanent scatterers are on the roofs of buildings.

elevation (Fig. 2). Parks and other areas of vegetation have few or no permanent scatterers.

Our subsidence rates represent a three-year average. Although subsidence may vary with time, the average range-change rate we obtain is very similar to recent estimates of subsidence from ground-based levelling^{2–4}, and we believe our estimates reflect rates over a longer term. The geography of subsidence also shows high rates in the areas of Lakeview (south shore of Lake Pontchartrain) and Kenner (near Louis Armstrong New Orleans International Airport), in east New Orleans and in regions bordering St Bernard Parish.

Lakeview is augmented with engineered fill, whereas the other areas are former wetlands, now drained and urbanized. The wetlands have highly organic soils; their subsidence results from drainage projects that promote soil desiccation, oxidation and compaction. In east New Orleans, subsidence rates have historically been the highest in south Louisiana, producing the lowest topography (3 to 5 m below sea level) and deepest flooding — the flooding after Hurricane Katrina caused a significant number of deaths. Subsidence here may also be increased by motion on the active Michoud fault².

Parts of St Bernard and Orleans Parishes west of Lake Borgne are experiencing subsidence rates of more than 20 mm yr^{-1} , including the levee system along the MRGO canal (Fig. 1, insets). Parts of this levee system were breached

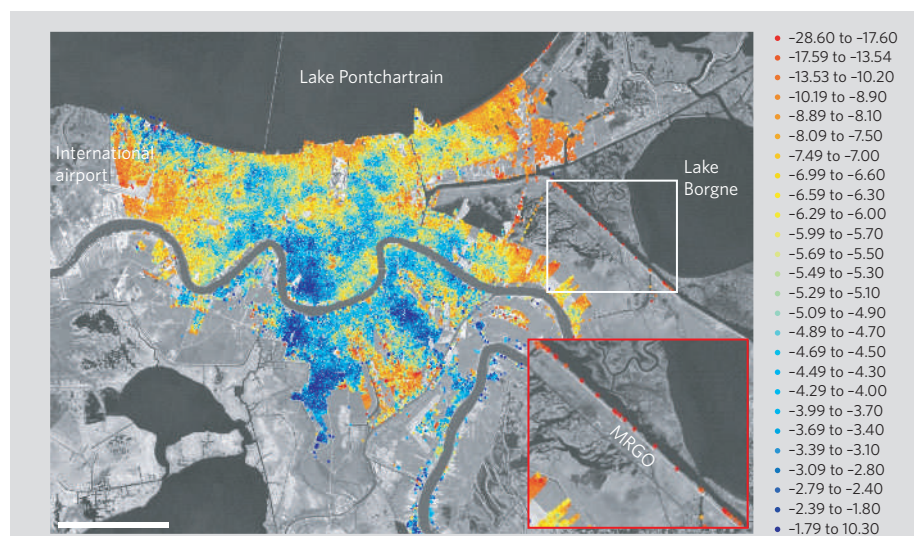


Figure 1 | Map showing rate of subsidence for permanent scatterers in New Orleans and vicinity during 2002–05. Velocity values are given in millimetres per year as range change in the direction of radar illumination. Negative values indicate motion away from the satellite, consistent with subsidence. International airport, Louis Armstrong New Orleans International Airport; MRGO, Mississippi River–Gulf Outlet canal. Insets show location (white frame) and magnified view (red frame) of the region west of Lake Borgne, including eastern St Bernard Parish. Note the high rates of subsidence on the levee bounding the MRGO canal. Large sections of the MRGO levee were breached when Hurricane Katrina struck on 29 August 2005 (see supplementary information). Scale bar, 10 km.

during the flooding associated with Hurricane Katrina, and this could be explained by the correlation we observe between the location of breach points and the high rate of subsidence beneath these levee sections (see supplementary information). Our subsidence estimates are probably minimum values considered over the lifetime of the levees, given that subsidence was most rapid in the first few years after their construction in the 1960s. Levee failure may have resulted from overtopping because the levees were too low — data collected after the storm⁵ indicate that water levels exceeded those expected by 0.9–1.7 m. Alternatively, the high subsidence rates we observe might reflect active faulting or a weak, easily compacted substrate,

promoting failure at or near the levee base. **Timothy H. Dixon***, **Falk Amelung***, **Alessandro Ferretti†**, **Fabrizio Novali‡**, **Fabio Rocca§**, **Roy Dokka§**, **Giovanni Sella||**, **Sang-Wan Kim***, **Shimon Wdowski***, **Dean Whitman¶**

*Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA

e-mail: tdixon@rsmas.miami.edu

†Tele-Rilevamento Europa, 20149 Milan, Italy

‡Dipartimento di Elettronica e Informazione, Politecnico di Milano, 20133 Milan, Italy

§Center for Geoinformatics, Louisiana State University, Baton Rouge, Louisiana 70803, USA

||National Geodetic Survey, Silver Spring, Maryland 20910-3282, USA

¶Department of Earth Sciences, Florida International University, Miami, Florida 33199, USA

1. Miller, L. & Douglas, B. C. *Nature* **248**, 407–409 (2004).
2. Dokka, R. K. *Geology* **34**, 281–284 (2006).
3. Shinkle, K. D. & Dokka, R. K. *NOAA Tech. Rep.* **50**, 135 (2004).
4. Burkett, V. R., Zilkoski, D. B. & Hart, D. in *US Geol. Surv. Subsidence Int. Grp. Conf., Proc. Tech. Meet., Galveston, Texas, Nov. 27–29, 2001* (eds Prince, K. R. & Galloway, D. L.) 63–70 (report 03-308, US Geol. Surv. Water Resour. Div., 2003).
5. *Performance Evaluation of the New Orleans and Southeast Louisiana Hurricane Protection System Rep. 2* (Interagency Performance Evaluation Task Force, 2006).

Supplementary information accompanies this communication on Nature's website.

Received 28 January; accepted 12 April 2006.

Competing financial interests: declared (see online communication).

doi:10.1038/441587a

SUPERFLUID HELIUM

Visualization of quantized vortices

When liquid helium is cooled to below its phase transition at 2.172 K, vortices appear with cores that are only ångströms in diameter, about which the fluid circulates with quantized angular momentum¹. Here we generate small particles of solid hydrogen that can be used to image the cores of quantized vortices in their three-dimensional environment of liquid helium. This technique enables the geometry and interactions of these vortices to be observed directly.

Since the discovery of quantized vortices², attempts have been made to visualize them. Although the ends of parallel vortices in an array could be located³, there has been no successful imaging of vortices in arbitrary three-dimensional configurations. Suspended particles can trace fluid motions, and the velocities of frozen particles in superfluid helium have been measured since hydrogen was first condensed for the purpose⁴. These particles usually have diameters larger than 10 µm^{5,6}, although smaller polymer particles have been used⁷.

Our technique generates smaller hydrogen particles by injecting a premixed gaseous solution of hydrogen, greatly diluted with helium, into liquid helium in its normal phase above the transition temperature (for details of methods, see supplementary information). This procedure yields a mist of randomly distributed hydrogen particles (Fig. 1a) that are smaller than the 2.7-µm resolution of our long-range microscope. The suspension so prepared is then cooled to below the transition temperature.

Images taken with a digital camera focused

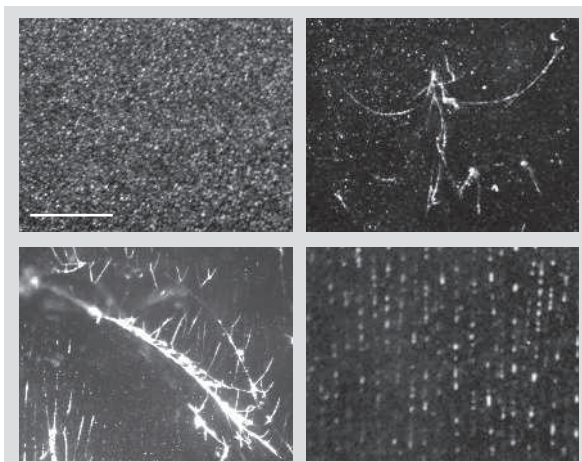


Figure 1 | Quantized vortex cores in liquid helium. **a–d**, Images of particles (light against dark background) obtained with a camera and 105-mm lens under different conditions: **a**, just above the transition temperature, when they are uniformly dispersed; **b**, **c**, on branching filaments at tens of millikelvin below the transition temperature; and **d**, regrouping along vertical lines for steady rotation about the vertical axis. In **b** and **c**, the particles on lines are evenly separated in small regions. Scale bar, 1 mm.

on a thin laser-illuminated sheet show that not only do the particles trace fluid motions, but a fraction of them also collect on slender filaments, which are often several millimetres long (Fig. 1b, c).

Particles respond in a complicated way to superfluid flows⁸ and can be trapped in vortex cores⁹. The following evidence suggests that the observed filaments are particles collected on such cores. First, these filaments appear only below the transition temperature. Second, when the liquid-helium cell is set in steady rotation, the particles arrange themselves along uniformly spaced lines (Fig. 1d). The lines are parallel to the axis of rotation, which is in the image plane. This observation agrees with the expectation that quantized vortices form a rectilinear array aligned with the axis¹. Third, if we

assume that our sheet illuminates a slice of such an array, we find that the number density of lines per unit area normal to the axis of rotation, for a series of rotation rates, is consistent with Feynman's rule¹⁰, which predicts about 2,000 Ω lines per cm², where Ω is the angular velocity of the container in radians per second.

The filaments are complex: for example, they may give rise to branched networks and to particles evenly spaced along lines. Others have speculated how particles could act as passive tracers of the flow^{7,8}; our images indicate that the presence of particles in the superfluid may transform the topology of vortex tangles by stabilizing forks in the vortices. This technique offers a tantalizing glimpse of new phenomena for future investigation.

Gregory P. Bewley†**, **Daniel P. Lathrop***, **Katepalli R. Sreenivasan*‡**

*Department of Physics, Institute for Research in Electronics and Applied Physics, Institute for Physical Sciences and Technology, University of Maryland, College Park, Maryland 20742, USA
e-mail: dpl@complex.umd.edu

†Department of Mechanical Engineering, Yale University, New Haven, Connecticut 06520, USA

‡International Centre for Theoretical Physics, Trieste 34014, Italy

1. Donnelly, R. J. *Quantized Vortices in Helium II* (Cambridge Univ. Press, Cambridge, 1991).
2. Onsager, L. *Proc. Int. Conf. Theor. Phys.* 877–880 (Science Council of Japan, Tokyo, 1953).
3. Yarmchuk, E. J., Gordon, M. J. V. & Packard, R. E. *Phys. Rev. Lett.* **43**, 214–218 (1979).
4. Chopra, K. L. & Brown, J. B. *Phys. Rev.* **108**, 157 (1957).
5. Chung, D. Y. & Critchlow, P. R. *Phys. Rev. Lett.* **14**, 892–894 (1965).
6. Murakami, M. & Ichikawa, N. *Cryogenics* **29**, 438–443 (1989).
7. Zhang, T. & Van Sciver, S. W. *Nature Phys.* **1**, 36–38 (2005).
8. Poole, D. R., Barengi, C. F., Sergeev, Y. A. & Vinen, W. F. *Phys. Rev. B* **71**, 064514–1–16 (2005).
9. Parks, P. E. & Donnelly, R. J. *Phys. Rev. Lett.* **16**, 45–48 (1966).
10. Feynman, R. P. in *Prog. Low Temp. Phys. I* (ed. Gorter, C. J.) 17–53 (North Holland, Amsterdam, 1955).

Supplementary information accompanies this communication on Nature's website.

Received 23 December 2005; accepted 28 April 2006.

Competing financial interests: declared none.

doi:10.1038/441588a

BRIEF COMMUNICATIONS ARISING online

▶ www.nature.com/bca see Nature contents.

PALAEOCEANOGRAPHY

Methane release in the Early Jurassic period

Arising from: D. B. Kemp, A. L. Coe, A. S. Cohen & L. Schwark *Nature* **437**, 396–399 (2005).

Dramatic global warming, triggered by release of methane from clathrates, has been postulated to have occurred during the early Toarcian age in the Early Jurassic period¹. Kemp *et al.*² claim that this methane was released at three points, as recorded by three sharp excursions of $\delta^{13}\text{C}_{\text{org}}$ of up to 3‰ magnitude. But they discount another explanation for the excursions: namely that some, perhaps all, of the rapid excursions could be a local signature of a euxinic basin caused by recycling of isotopically light carbon from the lower water column. This idea has been proposed previously (see ref. 3, for example) and is supported by the lack evidence for negative $\delta^{13}\text{C}$ excursions in coeval belemnite rostra⁴. Kemp *et al.* dismiss this alternative, claiming that each abrupt shift would have required the recycling of about double the amount of organic carbon that is currently present in the modern ocean; however, their measurements are not from an ocean but from a restricted, epicontinental seaway and so would not require whole-ocean mixing to achieve the excursions.

We also question aspects of the timing, magnitude and inferred consequences of the gas-hydrate release events. Kemp *et al.*² erect a cyclostratigraphic scheme on which they found their claim that the three carbon isotope spikes all occurred within a 60,000-year period. But they do not explain how a 3‰ recovery could have occurred between their second and third excursions. This interval is only 25% of the residence time of carbon in the oceans (180,000 yr; ref. 5) and is thus too short to represent a relaxation process. Their data show further very rapid declines in $\delta^{13}\text{C}_{\text{org}}$ values (for example, at 1.6 m and 2.9 m in their section) and these were not given equal status as gas-hydrate-release events. Furthermore, with three, and possibly five, closely spaced release events, there would not have been sufficient time to replenish the clathrate reservoirs between events.

Kemp *et al.*² note that one consequence of the release of huge amounts of methane (at least 5,000 gigatonnes of carbon by their calculations) was the development of a severe greenhouse climate and, as a direct (but unexplained) result, the extinction of many marine species. However, this is not supported by the work they cite. The authors suggest that there could have been a sudden rise in seawater temperatures coincident with methane release⁶, where “sudden” indicates a period of 0.6 to 0.7 million years (ref. 6). Note also that the temperature rise began in the *Dactylioceras tenuicostatum* subzone⁶, whereas the lowest $\delta^{13}\text{C}_{\text{org}}$ excursion of Kemp *et al.*² is not devel-

oped until late in the succeeding *Dactylioceras semicelatum* subzone. Furthermore, the younger two release events coincide with a proposed phase of marked cooling, according to stomatal index data⁷.

The claim by Kemp *et al.* that the first two methane-release events led to a 67% and 50% loss in species was derived from the work of Harries and Little⁸; however, the resolution of that study is too low to equate the extinction losses to the level of the excursions. The original data set⁹ and our later findings¹⁰ show that extinction in these Yorkshire coast sections occurs in Bed 31, where 12 of 14 benthic species go extinct. This crisis was 60,000 years before the first methane-release event, according to the timescale of Kemp *et al.* Other extinctions in the European region were even earlier, with many taxa disappearing at the Pliensbachian–Toarcian stage boundary¹¹. There is therefore no evidence to relate the Toarcian mass extinction to the effects of gas-hydrate dissociation events, although they do relate to the development of bottom-water anoxia.

Paul B. Wignall*, John M. McArthur†, Crispin T. S. Little*, Anthony Hallam‡

*School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK

e-mail: wignall@earth.leeds.ac.uk

†Department of Geological Sciences, University College London, London WC1E 6BT, UK

‡School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK

1. Hesselbo, S. P. *et al.* *Nature* **406**, 392–395 (2000).
2. Kemp, D. B., Coe, A. L., Cohen, A. S. & Schwark, L. *Nature* **437**, 396–399 (2005).
3. Sælen, G., Tyson, R. V., Talbot, M. R. & Telnæs, N. *Geology* **26**, 747–750 (1998).
4. van de Schootbrugge, B. *et al.* *Paleoceanography* **20**, PA3008, doi:10.1029/2004PA001102 (2005).
5. Katz, M. E. *et al.* *Mar. Geol.* **217**, 323–338 (2005).
6. Bailey, T. R., Rosenthal, Y., McArthur, J. M., van de Schootbrugge, B. & Thirlwall, M. F. *Earth Planet. Sci. Lett.* **212**, 307–320 (2003).
7. McElwain, J. C., Wade-Murphy, J. & Hesselbo, S. P. *Nature* **435**, 479–482 (2005).
8. Harries, P. J. & Little, C. T. S. *Paleogeog. Paleoclim. Paleogeol.* **154**, 39–66 (1999).
9. Little, C. T. S. Thesis, Univ. Bristol (1994).
10. Wignall, P. B., Newton, R. J. & Little, C. T. S. *Am. J. Sci.* **305**, 1014–1032 (2005).
11. Little, C. T. S. & Benton, M. J. *Geology* **23**, 495–498 (1995).

doi:10.1038/nature04905

PALAEOCEANOGRAPHY

Kemp et al. reply

Replying to: P. B. Wignall, J. M. McArthur, C. T. S. Little & A. Hallam *Nature* **442**, doi:10.1038/nature04905 (2006)

Wignall *et al.*¹ suggest that the abrupt negative $\delta^{13}\text{C}_{\text{org}}$ shifts that we recognize in our data set² could be localized features caused by recycling of isotopically light marine carbon. This overlooks two important lines of evidence²: that the early Toarcian negative $\delta^{13}\text{C}$ excursion is present in several sections around the world^{3–5}, and that the excursion is recorded in terrestrial plant material⁴. These crucial observations together preclude the possibility that the event was restricted to a localized, marine environment.

Wignall *et al.* cite evidence⁶ for a lack of a negative carbon-isotope excursion based on belemnite measurements from Yorkshire: however, this consists of just 24 scattered data points (we used 449 data points²) and the sample positions indicate that our abrupt $\delta^{13}\text{C}_{\text{org}}$ shifts were not sampled. Unambiguous evidence is also presented⁶ from a German section for a -2% shift in $\delta^{13}\text{C}_{\text{belemnite}}$. This shift coincides with a negative shift in $\delta^{13}\text{C}_{\text{org}}$ from the same section and correlates with our data. Our

original statement — that the recycling theory places impossible demands upon the marine carbon reservoir — is thus fully supported.

Establishing the cause of the roughly 3‰ recovery between the second and third abrupt negative $\delta^{13}\text{C}_{\text{org}}$ shifts was beyond the scope of our report², but we did not suggest that this recovery was the result of a ‘relaxation’ process. We also did not claim that hydrate reservoirs replenish themselves over short timescales¹. It is possible that each pulse of methane release represents dissociation of geographically distinct reservoirs. Wignall *et al.* question why other sharp negative $\delta^{13}\text{C}_{\text{org}}$ shifts are not ascribed to methane release, but we did in fact mention this possibility in our report².

The warming rate of sea water described by Bailey *et al.*⁷ is much slower than we suggested because their interpretation is based on a timescale that is incorrect, as shown by mathematically robust cyclostratigraphy². The inception of gradual global warming in the *tenuicostatum* subzone supports the theory that

warming began from volcanogenic sources^{2,4}. Nevertheless, the major change in temperature occurred in the *semicelatum* subzone².

The assumption by Wignall *et al.* that the two younger methane release events coincided with global cooling⁸ is incorrect: the correlation between the Bornholm section⁸ used to derive the stomatal index results and the Yorkshire section is based on low-resolution $\delta^{13}\text{C}$ data⁴. There is a discrepancy between this correlation⁸ and that provided by the dinoflagellate zones^{9,10} that prevents accurate correlation with our high-resolution data. The decrease in CO_2 and inferred cooling⁸ could have occurred between methane pulses as a result of increased weathering and associated CO_2 drawdown, which are known features of the event¹¹.

The faunal data of Harries and Little¹² reveal two distinct extinction horizons that are coincident with our first two $\delta^{13}\text{C}_{\text{org}}$ shifts, if their stratigraphy is used. However, even if this work has been superseded since our report², it has no bearing on the theory that methane-hydrate dissociation occurred during the early Toarcian, because faunal extinction was not an argument we used to support our theory.

D. B. Kemp*, A. L. Coe*, A. S. Cohen*, L. Schwark†

*Department of Earth Sciences, Centre for Earth, Planetary, Space & Astronomical Research, The Open University, Milton Keynes, MK7 6AA, UK
e-mail: d.kemp@open.ac.uk

†Geologisches Institut, Universität zu Köln, Köln 50674, Germany

1. Wignall, P. B., McArthur, J. M., Little, C. T. S. & Hallam, A.

Nature **442**, doi:10.1038/nature04905 (2006).

2. Kemp, D. B., Coe, A. L., Cohen, A. S. & Schwark, L. *Nature* **437**, 396–399 (2005).
3. Jenkyns, H. C., Gröcke, D. R. & Hesselbo, S. P. *Paleoceanography* **16**, 593–603 (2001).
4. Hesselbo, S. P. *et al.* *Nature* **406**, 392–395 (2000).
5. Gröcke, D. R., Hori, R. S. & Arthur, M. A. *EOS Trans.* **84**, 905 (2003).
6. van de Schootbrugge, B. *et al.* *Paleoceanography* **20**, PA3008, doi:10.1029/2004PA001102 (2005).
7. Bailey, T. R. *et al.* *Earth Planet. Sci. Lett.* **212**, 307–320 (2003).
8. McElwain, J. C., Wade-Murphy, J. & Hesselbo, S. P. *Nature* **435**, 479–482 (2005).
9. Bucefalo-Palliani, R., Mattioli, E. & Riding, J. B. *Mar. Micropaleontol.* **46**, 223–245 (2002).
10. Koppelhus, E. B. & Nielsen, L. H. *Palynology* **18**, 139–194 (1994).
11. Cohen, A. S., Coe, A. L., Harding, S. M. & Schwark, L. *Geology* **32**, 157–160 (2004).
12. Harries, P. J. & Little, C. T. S. *Palaogeog. Palaeoclimatol. Palaeoecol.* **154**, 39–66 (1999).

doi:10.1038/nature04906

A putative flip-flop switch for control of REM sleep

Jun Lu¹, David Sherman¹, Marshall Devor^{1,2} & Clifford B. Saper¹

Rapid eye movement (REM) sleep consists of a dreaming state in which there is activation of the cortical and hippocampal electroencephalogram (EEG), rapid eye movements, and loss of muscle tone. Although REM sleep was discovered more than 50 years ago, the neuronal circuits responsible for switching between REM and non-REM (NREM) sleep remain poorly understood. Here we propose a brainstem flip-flop switch, consisting of mutually inhibitory REM-off and REM-on areas in the mesopontine tegmentum. Each side contains GABA (γ -aminobutyric acid)-ergic neurons that heavily innervate the other. The REM-on area also contains two populations of glutamatergic neurons. One set projects to the basal forebrain and regulates EEG components of REM sleep, whereas the other projects to the medulla and spinal cord and regulates atonia during REM sleep. The mutually inhibitory interactions of the REM-on and REM-off areas may form a flip-flop switch that sharpens state transitions and makes them vulnerable to sudden, unwanted transitions—for example, in narcolepsy.

Previous models for the regulation of REM sleep¹ have emphasized the interactions of brainstem cholinergic neurons, which fire most rapidly during REM sleep^{2,3}, and brainstem monoaminergic neurons, which cease firing during REM sleep^{4–6}. There is pharmacological evidence that cholinergic stimulation promotes REM sleep^{7,8}, whereas monoamine re-uptake inhibitors, which increase extracellular levels of serotonin and noradrenaline, inhibit REM sleep⁹. However, selective lesions of either the cholinergic or monoaminergic nuclei in the brainstem have relatively limited effects on REM sleep^{10–12}, suggesting that there must be additional circuitry that regulates switching into and out of REM sleep.

To identify the switching circuitry for REM sleep, we traced the sites of convergence of two descending pathways from the hypothalamus that have recently been found to control REM sleep in rats. The orexin (hypocretin) neurons in the lateral hypothalamus are excitatory and cease firing during REM sleep^{13–15}. Their loss causes narcolepsy, in which fragments of REM sleep—including dreaming (hypnagogic hallucinations) and muscle atonia (cataplexy)—can occur against the background of wakefulness^{13,16}. Thus, we hypothesized that the orexin neurons excite a REM-off neuron population when they are active, and immunohistochemically examined their outputs in rats. The extended part of the ventrolateral preoptic nucleus (eVLPO) contains REM-active neurons that contain the inhibitory neurotransmitters galanin and GABA¹⁷. Thus, the eVLPO probably inhibits a REM-off site, and we traced its output with an anterograde tracer, adeno-associated virus expressing the gene for green fluorescent protein (AAV-GFP) ($n = 3$). These two pathways converged in a crescent-shaped arc of tissue in the mesopontine tegmentum, consisting of the ventrolateral part of the periaqueductal grey matter (vlPAG), at the opening of the fourth ventricle and sweeping laterally into the lateral pontine tegmentum (LPT) (Fig. 1a–e). This arc of tissue also contains many neurons that express the orexin 2 receptor (hypocretin receptor 2)¹⁸—the absence of which causes narcolepsy in dogs and mice.

Identifying a REM-off region

To determine the role of this region in regulating REM sleep, we

placed cell-specific lesions at these sites using the excitotoxin ibotenic acid (Fig. 1f, g). Consistent with an earlier report of short-term enhancement of REM by muscimol injection into this region in cats¹⁹, we found that complete lesions (>90% neuron loss) of either the vlPAG or the LPT doubled the amount of REM sleep, especially during the dark period (Table 1), and increased both the number and duration of bouts of REM sleep at night (Table 1). LPT, but not vlPAG, lesions increased REM bout number during the light period (Fig. 1h, i), as well as causing occasional episodes of sleep-onset REM (Fig. 1j), and 2–3 attacks per night of a cataplexy-like state with desynchronized EEG and atonia lasting 1–4 min and intruding between waking behaviours. Occasional episodes of atonia without theta EEG lasting 1–3 min also occurred immediately after an NREM sleep episode (similar to sleep-offset paralysis in humans with narcolepsy). In contrast, control lesions of the dorsal (DRN) and median (MnR) raphe nuclei (with 80% loss of serotonergic neurons) by 5,7-dihydroxytryptamine (5,7-DHT) just medial to and sparing the vlPAG, or ibotenic acid lesions of the medial pontine reticular formation (MPRF) just medial to and sparing the LPT, did not produce any significant effects on sleep–wake behaviour (Table 1).

Identifying a REM-on region

To determine whether this REM-off region might inhibit a REM-on site, we combined anterograde tracing with examination of c-Fos protein expression during enhanced REM sleep ($n = 32$). The REM-off region provided intense projections to an adjacent region of mesopontine tegmentum, which showed high levels of c-Fos expression in enhanced REM states (Fig. 2a), and included the sublaterodorsal nucleus (SLD)²⁰ (equivalent to the subcoeruleus area or peri-locus coeruleus alpha in cats²¹) and the periventricular grey matter, including a dorsal extension of the SLD and the precoeruleus (PC) region. This pattern is consistent with earlier reports that the neurons in the SLD are REM-active and that disinhibition of the SLD with bicuculline or stimulation by kainic acid increases REM-sleep-like behaviour, whereas muscimol in this site inhibits REM sleep^{20–22}.

¹Department of Neurology and Program in Neuroscience, Harvard Medical School and Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. ²Department of Cell and Animal Biology, Hebrew University, Jerusalem 91904, Israel.

Interactions between REM-on and REM-off areas

Next, we explored the relationship between these putative REM-on and REM-off neurons. Injection of CTB (cholera toxin subunit B) into the vlPAG ($n = 3$; Fig. 2b) or LPT ($n = 5$) produced a large number of retrogradely labelled neurons in the dorsal and ventral SLD (vSLD), and about half of these also contained messenger RNA encoding the glutamate decarboxylase 67 kDa isoform (*Gad67*; also known as *Gad1*; Fig. 2c, g) (whereas no GABAergic neurons were found in the PC). Injection of the anterograde tracer AAV-GFP into the SLD ($n = 2$) extensively labelled axon terminals within the region of the vlPAG and LPT containing GABAergic neurons (Fig. 2d). A recent study had shown that many of the inputs to the SLD from the vlPAG and LPT arise from GABAergic neurons²³. Thus, the SLD GABAergic REM-on neurons may inhibit GABAergic REM-off neurons in the vlPAG–LPT, and *vice versa*. This circuit arrangement, which allows mutual inhibition between REM-on and REM-off neurons, suggests a flip–flop switch arrangement in which each side—by inhibiting the other—also disinhibits (and thus reinforces) its own firing (Fig. 3). A flip–flop switch has the property of being stable in either one end state or the other, but avoiding intermediate states^{24,25}. This type of switch causes sharp transitions, which are typical of the rapid switching from NREM sleep to REM sleep and back again.

To test the role of the REM-on GABAergic neurons in the SLD, we placed ibotenic acid lesions including the SLD neurons both ventrally and in the more dorsal periventricular grey matter (Fig. 2e). Lesions

causing >90% cell loss in both components of the SLD produced animals in which REM sleep was markedly diminished and individual bouts were very short (Fig. 2f), but there was a marked increase in NREM–REM transitions. In some transitions between NREM sleep and REM sleep, fragments of REM sleep (theta EEG but with hyperactive muscle tone; that is, REM without atonia) would occur, but these were almost immediately abolished and merged into waking behaviours. The numbers of transitions during the light period doubled (REM bout number = 54.7 ± 5.6 (mean $\pm$ s.e.m.)) compared with control animals (REM bout number = 28.3 ± 3.5), but the duration of these presumed REM epochs was so short (lesioned 0.35 ± 0.1 min versus control 2.2 ± 0.2 min) that the animals effectively had about 30% of the normal amount of daytime REM sleep, and 24-h REM sleep was reduced to approximately 40% of control levels (Table 1). Thus, a tendency towards more frequent transitions was seen when lesions were placed into either the vlPAG–LPT or the SLD—a key property of flip–flop switches^{24,25}.

In four cases in which the lesions also included the PC, we could not detect any conventional REM sleep segments. However, we did observe very large numbers of transitions between NREM and waking, suggesting that attempts to transition into REM episodes instead may have proceeded into wakefulness. Thus, integrity of the REM-on region that we have identified is necessary for REM sleep to occur. In contrast, ibotenic acid lesions of areas that surround the SLD, such as the cholinergic pedunculopontine tegmental

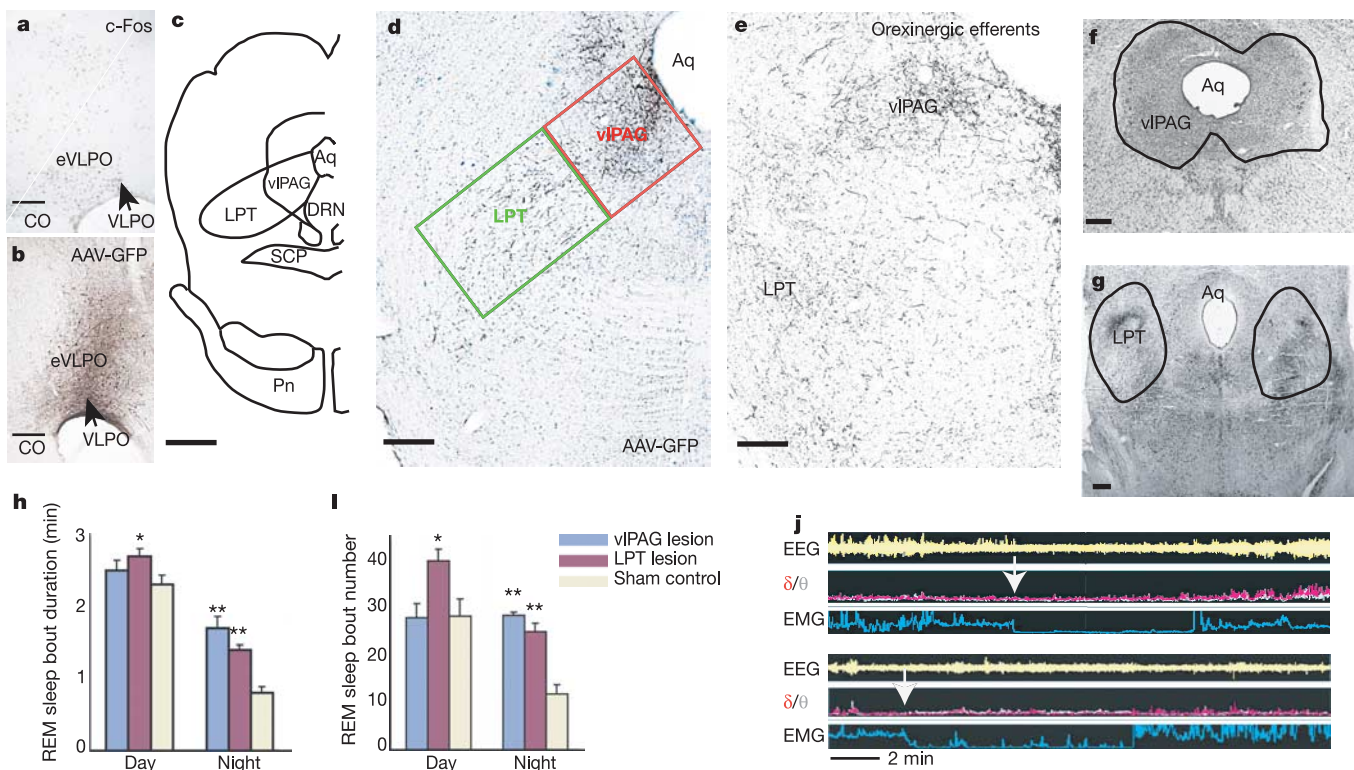


Figure 1 | The ventrolateral periaqueductal grey matter (vlPAG) and the lateral pontine tegmentum (LPT) constitute REM-off areas. **a**, During sleep, c-Fos-immunoreactive (brown) neurons are seen in the ventrolateral preoptic nucleus (VLPO) cluster (arrow) and extended VLPO (eVLPO), which preferentially cause NREM and REM sleep, respectively. **b–d**, After filling this region with the anterograde tracer AAV-GFP (**b**), descending axons ramify extensively in the vlPAG and LPT (**c**), as shown by black axons against a blue Nissl-stained background (**d**). Red and green boxes in **d** are the counting boxes for the vlPAG and the LPT, respectively. **e**, Projections from the orexin neurons, which inhibit REM sleep, are immunohistochemically stained black and are virtually superimposable with the VLPO projection.

f–i, Bilateral lesions in the vlPAG (**f**) or in the LPT (**g**) double the amount of REM sleep. With vlPAG lesions, REM increases mainly in the night, owing to an increase in both bout length (**h**) and number (**i**). Error bars are s.e.m. LPT lesions also increase REM sleep during the day, but mainly by increasing bout number. **j**, Lesions of the LPT also produce a cataplexy-like state (upper panel, arrow shows an episode of atonia in the EMG (electromyogram) during a wakeful period in the EEG (electroencephalogram)) and REM-onset sleep episodes (lower panel, arrow shows transition from wakefulness to REM sleep). Scale bars: **a**, **b**, 400 μ m; **c**, 1 mm; **d**, 400 μ m; **e**, 200 μ m; **f**, **g**, 400 μ m. Aq, cerebral aqueduct; CO, optic chiasm; DRN, dorsal raphe nuclei; Pn, pontine nuclei; SCP, superior cerebellar peduncle.

Table 1 | The effects of neurotoxic lesions in the pontine tegmentum on sleep

Lesion	NREM sleep	REM sleep	Wakefulness	Transitions per hour
Control (n = 8)	42.0 ± 2.7%	5.9 ± 0.6%	51.1 ± 3.7%	7.2 ± 0.9
vIPAG (n = 12)	40.4 ± 2.6%	11.0 ± 1.1%** (†86%)	48.6 ± 4.1%	9.4 ± 0.3* (†31%)
LPT (n = 9)	43.7 ± 2.9%	11.5 ± 0.5%** (†95%)	44.8 ± 3.9%	9.2 ± 0.5* (†28%)
LDT (n = 4)	42.9 ± 3.5%	6.2 ± 1.4%	50.7 ± 5.2%	10.2 ± 1.7* (†42%)
PPT (n = 6)	53.3 ± 2.5%* (†27%)	10.7 ± 0.7%* (†81%)	36.0 ± 3.6%* (†30%)	8.3 ± 1.43
vSLD (n = 5)	53.9 ± 4.6%* (†28%)	7.8 ± 1.0%* (†32%)	38.3 ± 6.0%* (†25%)	7.2 ± 1.3
SLD (n = 7)	40.7 ± 2.3%	2.3 ± 0.5%** (†61%)	57.0 ± 3.1%	19.1 ± 2.6** (†165%)
MPB (n = 4)	51.6 ± 3.6%* (†23%)	7.9 ± 1.3%* (†34%)	40.5 ± 3.4%* (†21%)	6.9 ± 0.4
DRN (n = 5)	42.1 ± 1.6%	5.9 ± 0.8%	52.0 ± 2.4%	7.1 ± 0.8
MnR (n = 6)	40.2 ± 3.6%	5.3 ± 1.6%	54.5 ± 5.9%	6.8 ± 0.6
MPRF (n = 8)	43.1 ± 2.5%	6.6 ± 0.7%	50.3 ± 3.5%	7.8 ± 0.8
LC (n = 6)	41.1 ± 3.7%	5.8 ± 1.9%	53.1 ± 6.3%	7.3 ± 0.4

Average percentage of sleep and wakefulness, and total transitions (per hour), across 24 h under a 12-h light-dark cycle were measured. Values are mean ± s.e.m. Abbreviations are defined in the text.
*P < 0.05 compared with controls.
**P < 0.01 compared with controls.

nucleus (PPT) (96.4 ± 3.9% loss of choline acetyltransferase (ChAT)-immunoreactive neurons) and the laterodorsal tegmental nucleus (LDT) (92.5 ± 3.1% loss of ChAT-immunoreactive neurons), or 6-hydroxydopamine lesions of the noradrenergic locus coeruleus (LC) (98.7 ± 0.3% loss of tyrosine hydroxylase (TH)-immunoreactive neurons) did not alter REM sleep (Table 1). Increased sleep after PPT lesions may have been due to including part of the medial parabrachial nucleus (MPB) (Table 1).

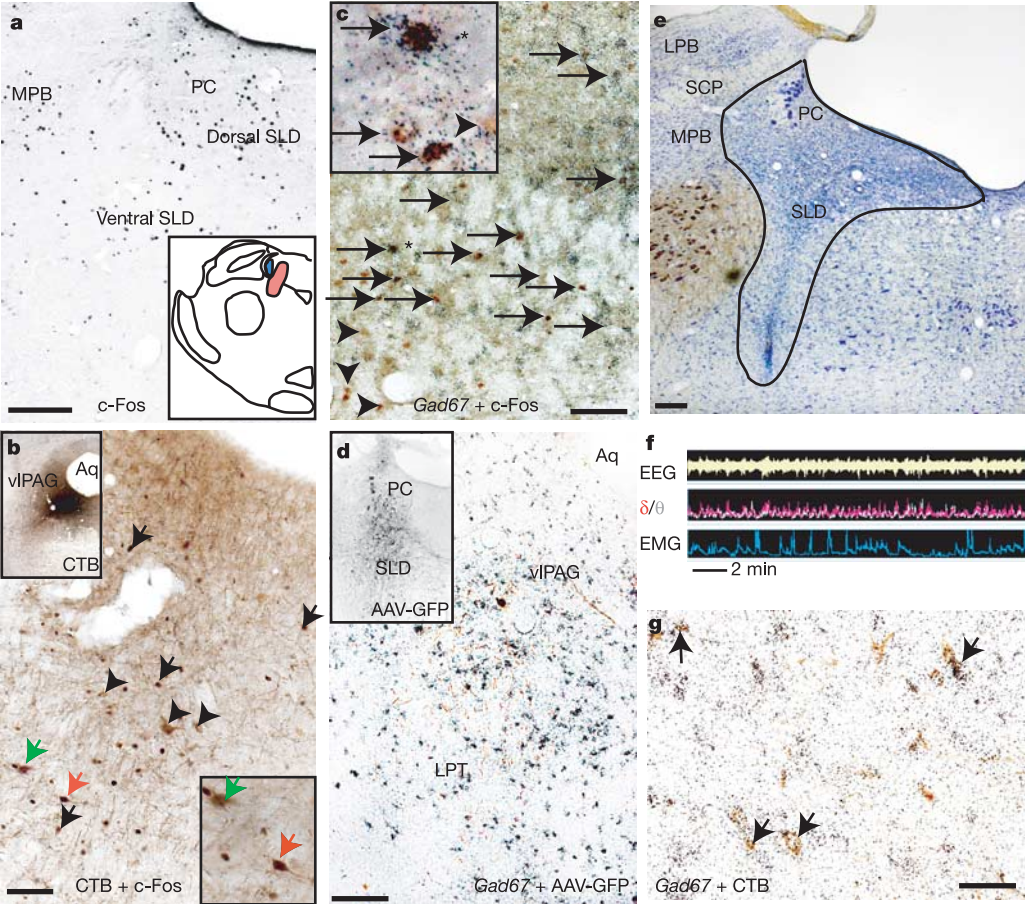


Figure 2 | The sublaterodorsal nucleus (SLD) and precoeruleus (PC) area constitute a REM-on region. **a**, During enhanced REM sleep (triggered by a 3-h dark exposure during the morning), c-Fos-immunoreactive neurons are seen in the SLD (orange in inset), PC (blue in inset) and the adjacent medial parabrachial nucleus (MPB). **b**, Many of these neurons in the SLD can be retrogradely labelled by injections of CTB into the vIPAG (injection site shown in upper inset; arrowheads and green arrows point to retrogradely labelled neurons; black and red arrows to double-labelled neurons; cells indicated by green and red arrows are shown in the bottom inset at a higher magnification). **c**, **g**, Most of the c-Fos-immunoreactive neurons (c; arrowheads indicate c-Fos-positive cells; arrows double-labelled

neurons) and retrogradely labelled neurons in the SLD (g; arrows indicate double-labelled cells) also contain *Gad67* mRNA and are presumably GABAergic. Asterisk indicates area shown at higher magnification in inset. **d**, Injection of the anterograde tracer AAV-GFP into the SLD (inset) demonstrates that this projection (brown) targets a population of neurons in the vIPAG and LPT that also contain *Gad67* mRNA (black silver grains). **e**, **f**, Lesions of the SLD and PC (e; black line indicates the limits of the lesion) greatly reduced REM sleep. Animals could not sustain REM periods, resulting in frequent transitions between NREM sleep, REM sleep and waking (f). Scale bars: **a**, 250 μm; **b**, 125 μm; **c**, 50 μm; **d**, 250 μm; **e**, 125 μm; **g**, 50 μm.

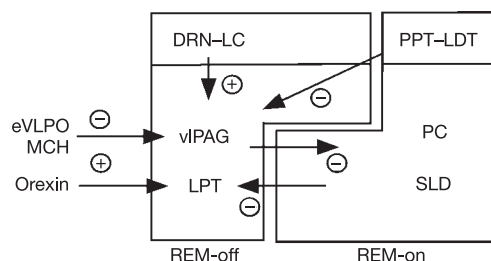


Figure 3 | The interrelationship of the two halves of the REM switch. The REM-off region is identified by the overlap of inputs from the orexin neurons and the eVLPO. These neurons in the vPAG and LPT have a mutually inhibitory interaction with REM-on GABAergic neurons of the vSLD, but also inhibit REM generator circuitry in the remainder of the SLD and the PC. Note that cholinergic neurons in the pedunculopontine and laterodorsal tegment nuclei (PPT–LDT) are REM-on and may inhibit the LPT (as cholinergic agonists injected in this region cause REM states), but are not directly inhibited by it, and thus are not part of the mutually inhibitory flip–flop switch. Similarly, serotonergic dorsal raphe and noradrenergic locus coeruleus (DRN–LC) neurons activate the REM-off circuitry, and thus monoamine re-uptake inhibitors, such as antidepressants, can dramatically suppress REM sleep. However, they also are not inhibited directly by the SLD, and hence are not part of the mutually inhibitory flip–flop switch.

Combined lesions of the SLD REM-on area and the LPT REM-off area ($n = 6$) produced the same effects as large SLD lesions (Table 1), suggesting that REM-off cells decrease REM sleep by inhibiting REM-on neurons. This flip–flop switch model, which is constructed from consideration of the underlying circuitry and selective lesion placement, will require further testing by examining neuronal firing states and their interactions across the wake–sleep

cycle. However, our maps provide a substrate for testing these theories.

Identifying pathways for REM atonia

We next examined the mechanisms by which the REM-on areas may produce REM sleep by placing cell-specific lesions with ibotenic acid selectively into different parts of the field. Consistent with earlier reports in cats and humans^{12,26–28}, ibotenic acid lesions of the vSLD (below the periventricular grey matter) alone caused loss of atonia in almost every REM sleep episode (Fig. 4a, b). Animals with vSLD lesions could not maintain atonia, showing muscle jerks and full lunging movements, and even some complex behaviours such as walking, apparently while in REM sleep. The small increase in total sleep (REM and NREM) following vSLD lesions appeared to be due to including part of the MPB in some lesions (Table 1).

We examined the projections from the SLD that mediate REM atonia by injecting it with the anterograde tracer AAV–GFP ($n = 8$; Fig. 4c). As in earlier reports, we traced axons through the medial pontine and medullary reticular formation to the spinal ventral horn²⁹. Earlier studies using electrolytic lesions had indicated that a relay through the medial medullary reticular formation has a crucial role in this pathway³⁰, so we placed large lesions in this site by using the cell-specific toxin orexin B–saporin (2 μ l of a 0.1% (w/v) solution; $n = 3$; Fig. 4d). However, consistent with a previous study on ibotenic acid lesions in cats³⁰, even extensive cell-specific lesions could not reduce REM atonia (Fig. 4e).

Therefore, we explored the direct spinal projection from the SLD. Retrograde labelling with Fluorogold (20 nl of a 5% (w/v) solution; $n = 14$) from the cervical spinal cord demonstrated large numbers of neurons in the vSLD, with an ipsilateral predominance (Fig. 4f), and about 10% of these neurons were c-Fos-positive during REM sleep enhanced by dark conditions ($n = 3$; Fig. 4g). Nearly all of the

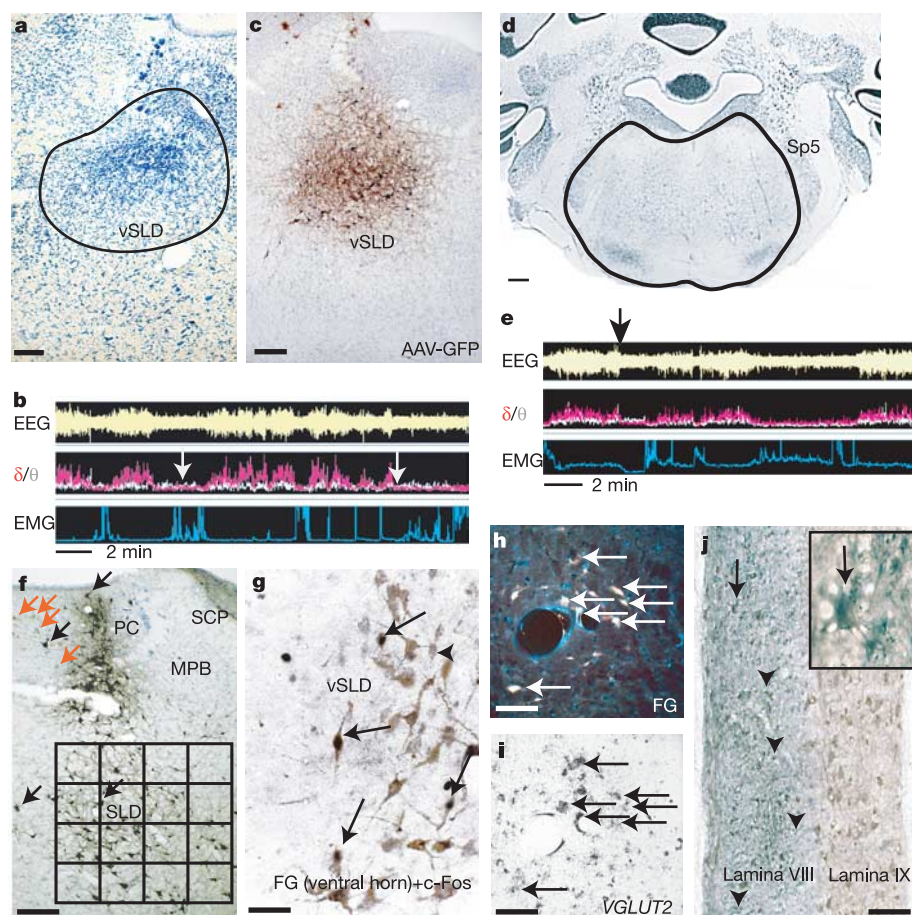


Figure 4 | The vSLD is responsible for REM atonia. **a, b**, Lesions of the vSLD (**a**) caused episodes of REM sleep (characterized by greater theta than delta power in the EEG; **b**) without atonia (arrows in **b**). **c**, Injections of the anterograde tracer AAV–GFP into the vSLD identified projections to the ventromedial medulla as well as to the spinal cord. **d, e**, However, animals with lesions of the ventromedial medulla with orexin B–saporin (**d**) had normal REM atonia (arrow in **e**). **f**, Retrograde labelling of the SLD spinal neurons with CTB (black; indicated by black arrows) showed that they were distinct from the cholinergic neurons (brown choline acetyltransferase immunostaining; indicated by red arrows). The grid in **f** indicates the SLD counting box in lesion experiments. **g–i**, About 10% of the SLD neurons that were retrogradely labelled by Fluorogold (FG) injection into the spinal cord were c-Fos-positive during enhanced REM sleep (arrows in **g**; arrowhead indicates neuron that is only retrogradely labelled), and nearly all of these retrogradely labelled neurons (white arrows in **h**) contained mRNA for *VGLUT2* (black arrows in **i**). **j**, SLD spinal terminals anterogradely labelled black with AAV–GFP (arrowheads) formed appositions with parvalbumin-immunoreactive lamina VIII neurons labelled blue–grey (see arrow in inset). In this horizontal spinal section, lamina IX motor neurons are stained brown for choline acetyltransferase. Scale bars: **a**, 200 μ m; **c**, 250 μ m; **d**, 300 μ m; **f**, 250 μ m; **g**, 100 μ m; **h, i**, 125 μ m; **j**, 150 μ m. Sp5, spinal trigeminal nucleus.

retrogradely labelled cells contained mRNA for vesicular glutamate transporter 2 (*VGLUT2*; also known as *Slc17a6*; Fig. 4h, i), but none was immunoreactive for choline acetyltransferase (Fig. 4f), indicating that the vSLD sends primarily glutamatergic, and therefore presumably excitatory, inputs to the spinal ventral horn. Antero-gradely labelled terminals from the SLD formed many appositions with parvalbumin-immunoreactive neurons in lamina VIII, most of which belong to the class of V1 interneurons that innervate motor neurons with terminals containing glycine, GABA or both^{31,32} (Fig. 4j). Thus, the neurons in the SLD may produce atonia by means of direct glutamatergic spinal projections—apparently not requiring a relay in the medial medulla—to interneurons that may inhibit spinal motor neurons by both glycinergic and GABAergic mechanisms³³.

Identifying pathways for REM EEG phenomena

Next, we explored the ascending projections from the REM-on region that might mediate the EEG changes associated with REM sleep. Hippocampal theta rhythm, which characterizes REM sleep in rats, is driven by septo-hippocampal projections, particularly inputs from medial septal GABAergic, but not cholinergic, neurons^{34,35}. After placing CTB injections into the medial septum ($n = 5$; Fig. 5a inset), retrogradely labelled cells in the REM-on area were mainly concentrated in the PC region (15.0 ± 1.6 cells per section per side), with a few (<5 cells per section per side) also scattered over the nearby periventricular grey matter and the MPB (Fig. 5a). Double-labelling for CTB and *VGLUT2* mRNA in these rats ($n = 7$) showed that the neurons with septal projections are primarily glutamatergic

(Fig. 5b). About 50% of these CTB-labelled neurons in the PC region were c-Fos-positive in association with enhanced REM sleep, that is, REM-active; $n = 5$; Fig. 5c). Ibotenic acid lesions targeting the PC region ($n = 4$) abolished EEG theta rhythm during sleep, so that presumptive REM epochs were marked by episodic periods of atonia and desynchronized EEG. There were no alterations in the total amounts of sleep or wakefulness. Thus, glutamatergic inputs from the REM-on PC region to the medial septum may have a key role in generating hippocampal theta EEG during REM sleep.

Conclusions

Our model for the regulation of REM sleep places the REM switch as part of a double flip–flip arrangement, as shown in Supplementary Fig. 1. In contrast to earlier models¹, which mainly emphasized interactions of cholinergic and monoaminergic neurons, our model hypothesizes that the main switching mechanism for REM sleep involves reciprocal interactions between GABAergic REM-off and REM-on populations. These mutually inhibitory neuronal populations can explain the sharp transitions into and out of REM sleep, as well as the increased frequency of transitions between REM and non-REM states when either the REM-on or REM-off population is weakened^{24,25}. Although the cholinergic and monoaminergic populations may still have an important modulatory role in this scheme, ablation of these cell groups in our experiments had minimal effects on REM sleep, so they most likely modulate the components of the main REM switch, but are not a part of it. Our results also demonstrate independent pathways mediating the atonia and EEG phenomena of REM, thus providing a basis for their occasional dissociation in pathological states. In an intact brain, our model places the REM flip–flip switch in a subsidiary relationship with respect to the wake–sleep flip–flip switch. This arrangement would prevent entry into REM states during wakefulness.

This model not only presents a set of testable hypotheses for the location of REM-on and REM-off elements in a flip–flip circuit, but also provides a context for understanding a variety of sleep disorders. In narcolepsy, the loss of excitatory orexin (hypocretin) neurons would weaken both the waking side of the wake–sleep flip–flip switch, as well as the REM-off side of the REM flip–flip switch. In particular, reduced activity in the LPT (where lesions caused a cataplexy-like state both in our rats and in humans²⁸) could produce attacks of cataplexy or sleep paralysis by allowing the activation of subsets of SLD REM-atonia neurons in the absence of sleep. Failure of the REM-off input to the PC may allow hypnagogic hallucinations in narcolepsy or peduncular hallucinosis in patients with midbrain stroke³⁶. Weakening the influence of the REM-off neurons would cause an increase in transitions both during wakefulness and sleep, which is characteristic of narcolepsy, and which is predicted by the flip–flip switch hypothesis.

On the other hand, damage to the SLD or its connections may cause REM sleep without atonia in our rats and in humans³⁷, and in patients with REM sleep behaviour disorder. This syndrome is often seen in patients with parkinsonism, sometimes years before the movement disorder becomes apparent. It has been shown that the earliest stages (stage I and II) of Parkinson's disease involve the lower brainstem including the subcoeruleus or SLD region³⁸. In summary, the flip–flip switch model for REM sleep may explain many of the key properties of REM sleep and many disorders of REM sleep in which its regulation is impaired. This model provides a framework for further work testing the physiological relationships of its components, and for examining their integrity in specific sleep disorders.

METHODS

Animals. Pathogen-free, adult male Sprague-Dawley rats (275–300 g; Harlan) were individually housed and had free access to food and water. All animals were housed under controlled conditions (12 h of light starting at 7:00; 100 lx) in an isolated ventilated chamber maintained at 20–22 °C. All protocols were

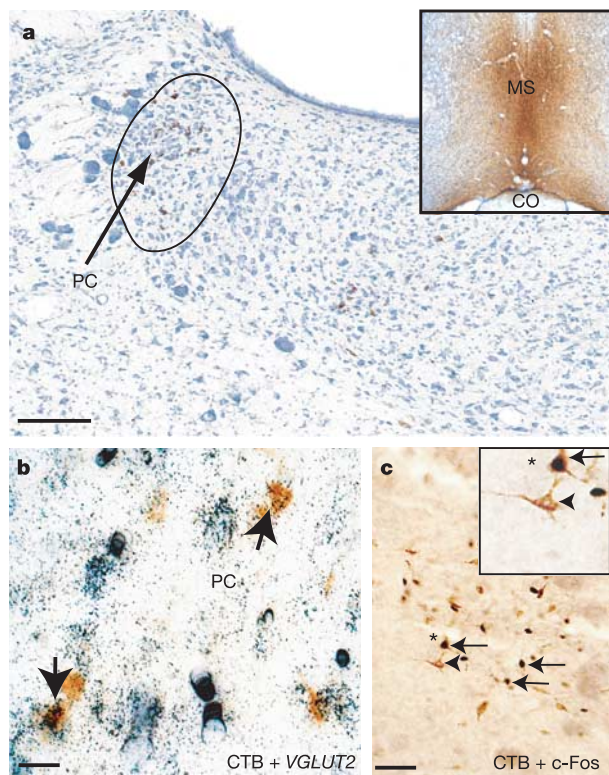


Figure 5 | PC neurons activate hippocampal theta during REM sleep. **a**, Many neurons in the PC (brown; arrow) are retrogradely labelled from the medial septum (MS; brown; CTB injection site in inset). **b**, Most of these neurons also contain *VGLUT2* mRNA (black silver grains), and are therefore glutamatergic. **c**, Many of the retrogradely labelled neurons are also c-Fos-immunoreactive during enhanced REM sleep. Double-labelled neurons are indicated by arrows and retrogradely labelled neurons by arrowheads. Asterisk indicates area of higher magnification in inset. Scale bars: **a**, 200 μ m; **b**, 20 μ m; **c**, 200 μ m.

approved by the Institutional Animal Care and Use Committees of Beth Israel Deaconess Medical Center and Harvard Medical School.

Injections of tracers and toxins. Injections of anterograde tracers (AAV-GFP, 50–250 nl), retrograde tracers (CTB, 3–10 nl of a 1% (w/v) solution; Fluorogold, 10–50 nl of a 5% (w/v) solution), or cell-specific toxins (ibotenic acid, 20–30 nl of a 10% (w/v) solution; orexin B–saporin, 10–20 nl; 6-hydroxydopamine, 10 nl of a 10% (w/v) solution; 5,7-dihydroxytryptamine, 50 nl of a 20% (w/v) solution) were made stereotactically under chloral hydrate anaesthesia (see Supplementary Methods).

EEG/EMG recording and sleep analysis. Under chloral hydrate anaesthesia, four EEG (electroencephalogram) screws were implanted into the skull, and two flexible EMG (electromyogram) wire electrodes were placed into the neck muscles. These were attached to a socket pedestal, which was glued to the skull with dental acrylic and the incision was closed (see Supplementary Methods). One week after surgery, the sockets were connected via flexible recording cables and a commutator to a Grass polygraph and signals were digitized by an Apple Macintosh computer running ICELUS software (M. Opp, University of Michigan). The EEG/EMG was recorded for 24 h at the end of the third week after surgery, and scored by standard criteria (see Supplementary Methods). REM sleep enhancement was produced by exposing animals to three hours of darkness during the first part of the light cycle, which increases REM sleep by 2–3-fold in albino rats¹⁷.

Histology, immunohistochemistry and *in situ* hybridization. At the end of each experiment, animals were deeply anaesthetized with chloral hydrate and perfused through the heart with 4% buffered paraformaldehyde. Brains were removed, equilibrated in 20% sucrose, and sectioned at 40 µm into four series. Immunohistochemistry for c-Fos, choline acetyltransferase, tyrosine hydroxylase, serotonin, cholera toxin B subunit, and green fluorescent protein were performed as described previously³⁹. (See Supplementary Methods for details.) Double-labelling was done by staining the reaction product from the first antibody black with cobalt-diaminobenzidine (DAB), and the second brown with DAB. In triple labelling, the third antibody was stained using the Vector SG grey reaction. *In situ* hybridization for *VGLUT2* and *Gad67* was performed as described previously³⁹. (See Supplementary Methods for details.)

Cell counts and statistics. Cell counts of remaining neurons at lesion sites were done by counting cells with a clearly outlined nucleus in Nissl or immunostaining, within a defined counting box as described previously¹⁷. Counting boxes were placed in predefined locations with respect to tissue landmarks and corrected by Abercrombie's factor (see Supplementary Methods). Statistical differences in cell counts and sleep behaviour were assessed with analysis of variance (ANOVA) followed by a *post-hoc* Student's *t*-test.

Received 20 November 2005; accepted 20 March 2006.

Published online 10 May 2006.

- McCarley, R. W. Mechanisms and models of REM sleep control. *Arch. Ital. Biol.* **142**, 429–467 (2004).
- Kayama, Y., Ohta, M. & Jodo, E. Firing of 'possibly' cholinergic neurons in the rat laterodorsal tegmental nucleus during sleep and wakefulness. *Brain Res.* **569**, 210–220 (1992).
- Steriade, M., Pare, D., Datta, S., Oakson, G. & Curro Dossi, R. Different cellular types in mesopontine cholinergic nuclei related to ponto-geniculo-occipital waves. *J. Neurosci.* **10**, 2560–2579 (1990).
- Wu, M. F. *et al.* Activity of dorsal raphe cells across the wake–sleep cycle and during cataplexy in narcoleptic dogs. *J. Physiol. (Lond.)* **554**, 202–215 (2004).
- Hobson, J. A., McCarley, R. W. & Nelson, J. P. Location and spike-train characteristics of cells in anterodorsal pons having selective decreases in firing rate during desynchronized sleep. *J. Neurophysiol.* **50**, 770–783 (1983).
- Aston-Jones, G. & Bloom, F. E. Activity of norepinephrine-containing locus coeruleus neurons in behaving rats anticipates fluctuations in the sleep–waking cycle. *J. Neurosci.* **1**, 876–886 (1981).
- Kubin, L. Carbachol models of REM sleep: recent developments and new directions. *Arch. Ital. Biol.* **139**, 147–168 (2001).
- Coleman, C. G., Lydic, R. & Baghdooyan, H. A. M2 muscarinic receptors in pontine reticular formation of C57BL/6J mouse contribute to rapid eye movement sleep generation. *Neuroscience* **126**, 821–830 (2004).
- Wilson, S. & Argyropoulos, S. Antidepressants and sleep: a qualitative review of the literature. *Drugs* **65**, 927–947 (2005).
- Jones, B. E., Harper, S. T. & Halaris, A. E. Effects of locus coeruleus lesions upon cerebral monoamine content, sleep–wakefulness states and the response to amphetamine in the cat. *Brain Res.* **124**, 473–496 (1977).
- Mouret, J. & Coindet, J. Polygraphic evidence against a critical role of the raphe nuclei in sleep in the rat. *Brain Res.* **186**, 273–287 (1980).
- Shouse, M. N. & Siegel, J. M. Pontine regulation of REM sleep components in cats: integrity of the pedunculopontine tegmentum (PPT) is important for phasic events but unnecessary for atonia during REM sleep. *Brain Res.* **571**, 50–63 (1992).
- Chemelli, R. M. *et al.* Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation. *Cell* **98**, 437–451 (1999).
- Lee, M. G., Hassani, O. K. & Jones, B. E. Discharge of identified orexin/hypocretin neurons across the wake–sleep cycle. *J. Neurosci.* **25**, 6716–6720 (2005).
- Mileykovskiy, B. Y., Kiyashchenko, L. I. & Siegel, J. M. Behavioral correlates of activity in identified hypocretin/orexin neurons. *Neuron* **46**, 787–798 (2005).
- Thannickal, T. C. *et al.* Reduced number of hypocretin neurons in human narcolepsy. *Neuron* **27**, 469–474 (2000).
- Lu, J. *et al.* Selective activation of the extended ventrolateral preoptic nucleus during rapid eye movement sleep. *J. Neurosci.* **22**, 4568–4576 (2002).
- Marcus, J. N. *et al.* Differential expression of orexin receptors 1 and 2 in the rat brain. *J. Comp. Neurol.* **435**, 6–25 (2001).
- Sastre, J. P., Buda, C., Kitahama, K. & Jouvet, M. Importance of the ventrolateral region of the periaqueductal gray and adjacent tegmentum in the control of paradoxical sleep as studied by muscimol microinjections in the cat. *Neuroscience* **74**, 415–426 (1996).
- Boissard, R. *et al.* The rat ponto-medullary network responsible for paradoxical sleep onset and maintenance: a combined microinjection and functional neuroanatomical study. *Eur. J. Neurosci.* **16**, 1959–1973 (2002).
- Sakai, K., Crochet, S. & Onoe, H. Pontine structures and mechanisms involved in the generation of paradoxical (REM) sleep. *Arch. Ital. Biol.* **139**, 93–107 (2001).
- Onoe, H. & Sakai, K. Kainate receptors: a novel mechanism in paradoxical (REM) sleep generation. *Neuroreport* **6**, 353–356 (1995).
- Boissard, R. *et al.* Localization of the GABAergic and non-GABAergic neurons projecting to the sublaterodorsal nucleus and potentially gating paradoxical sleep onset. *Eur. J. Neurosci.* **18**, 1627–1639 (2003).
- Saper, C. B., Chou, T. C. & Scammell, T. E. The sleep switch: hypothalamic control of sleep and wakefulness. *Trends Neurosci.* **24**, 726–731 (2001).
- Saper, C. B., Scammell, T. E. & Lu, J. Hypothalamic regulation of sleep and circadian rhythms. *Nature* **437**, 1257–1263 (2005).
- Sastre, J. P. & Jouvet, M. Oneiric behavior in cats. *Physiol. Behav.* **22**, 979–989 (1979).
- Sanford, L. D. *et al.* Sleep patterning and behaviour in cats with pontine lesions creating REM without atonia. *J. Sleep Res.* **3**, 233–240 (1994).
- Plazzi, G. *et al.* Pontine lesions in idiopathic narcolepsy. *Neurology* **46**, 1250–1254 (1996).
- Rye, D. B. *et al.* Medullary and spinal efferents of the pedunculopontine tegmental nucleus and adjacent mesopontine tegmentum in the rat. *J. Comp. Neurol.* **269**, 315–341 (1988).
- Sastre, J. P., Sakai, K. & Jouvet, M. Are the gigantocellular tegmental field neurons responsible for paradoxical sleep? *Brain Res.* **229**, 147–161 (1981).
- Alvarez, F. J. *et al.* Postnatal phenotype and localization of spinal cord V1 derived interneurons. *J. Comp. Neurol.* **493**, 177–192 (2005).
- Taal, W. & Holstege, J. C. GABA and glycine frequently colocalize in terminals on cat spinal motoneurons. *Neuroreport* **5**, 2225–2228 (1994).
- Chase, M. H., Soja, P. J. & Morales, F. R. Evidence that glycine mediates the postsynaptic potentials that inhibit lumbar motoneurons during the atonia of active sleep. *J. Neurosci.* **9**, 743–751 (1989).
- Kapas, L. *et al.* The effects of immunolesions of nerve growth factor-receptive neurons by 192 IgG-saporin on sleep. *Brain Res.* **712**, 53–59 (1996).
- Gerashchenko, D., Salin-Pascual, R. & Shiromani, P. J. Effects of hypocretin-saporin injections into the medial septum on sleep and hippocampal theta. *Brain Res.* **913**, 106–115 (2001).
- Manford, M. & Andermann, F. Complex visual hallucinations. Clinical and neurobiological insights. *Brain* **121**, 1819–1840 (1998).
- Kimura, K. *et al.* A discrete pontine ischemic lesion could cause REM sleep behavior disorder. *Neurology* **55**, 894–895 (2000).
- Braak, H. *et al.* Parkinson's disease: affection of brain stem nuclei controlling premotor and motor neurons of the somatomotor system. *Acta Neuropathol. (Berl.)* **99**, 489–495 (2000).
- Chou, T. C. *et al.* Critical role of dorsomedial hypothalamic nucleus in a wide range of behavioral circadian rhythms. *J. Neurosci.* **23**, 10691–10702 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Q. H. Ha and M. Ha for technical expertise. This work was supported by United States Public Health Service grants.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.L. (jlu@bidmc.harvard.edu) or C.B.S. (csaper@bidmc.harvard.edu).

NFAT dysregulation by increased dosage of *DSCR1* and *DYRK1A* on chromosome 21

Joseph R. Arron^{1*}, Monte M. Winslow^{2*}, Alberto Polleri^{1*}, Ching-Pin Chang³, Hai Wu¹, Xin Gao¹, Joel R. Neilson², Lei Chen¹, Jeremy J. Heit⁴, Seung K. Kim⁴, Nobuyuki Yamasaki⁷, Tsuyoshi Miyakawa⁷, Uta Francke⁵, Isabella A. Graef^{1*} & Gerald R. Crabtree^{1,4,6}

Trisomy 21 results in Down's syndrome, but little is known about how a 1.5-fold increase in gene dosage produces the pleiotropic phenotypes of Down's syndrome. Here we report that two genes, *DSCR1* and *DYRK1A*, lie within the critical region of human chromosome 21 and act synergistically to prevent nuclear occupancy of NFATc transcription factors, which are regulators of vertebrate development. We use mathematical modelling to predict that autoregulation within the pathway accentuates the effects of trisomy of *DSCR1* and *DYRK1A*, leading to failure to activate NFATc target genes under specific conditions. Our observations of calcineurin- and *Nfatc*-deficient mice, *Dscr1*- and *Dyrk1a*-overexpressing mice, mouse models of Down's syndrome and human trisomy 21 are consistent with these predictions. We suggest that the 1.5-fold increase in dosage of *DSCR1* and *DYRK1A* cooperatively destabilizes a regulatory circuit, leading to reduced NFATc activity and many of the features of Down's syndrome. More generally, these observations suggest that the destabilization of regulatory circuits can underlie human disease.

Genetic regulatory circuits evolve to be well-suited for their biological tasks and to be robust under the varying conditions encountered over the course of development and during responses to environmental stimuli^{1,2}. However, the operation of feedback loops, buffers and amplifiers within these circuits could also make them susceptible to conditions other than those that drove their evolution. One such condition would be chromosomal trisomy, the best-known example of which is Down's syndrome. The many features of Down's syndrome include neurological, skeletal, cardiovascular and immunological defects, and are generally thought to originate from a 1.5-fold increase in the dosage of genes within a critical region of chromosome 21, which is present in triplicate in all cases of Down's syndrome^{3–6}. In general, chromosomal trisomy is lethal, suggesting that the buffering mechanisms of many genetic regulatory circuits are susceptible to variations in gene dosage.

When studying the developmental roles of calcineurin and NFAT signalling, one of the authors (I.A.G.) noted striking similarities between the phenotypic features of Down's syndrome and mice carrying deletions of genes encoding components of the NFAT signalling pathway. This pathway, which is a critical regulator of vertebrate development and organogenesis^{7,8}, is initiated by Ca²⁺ entry and results in calcineurin activation. Calcineurin dephosphorylates NFATc proteins, leading to their nuclear entry and assembly with partner proteins (NFATn) to form NFAT transcription complexes⁹. Rephosphorylation by an unknown priming kinase and glycogen synthase kinase 3 (GSK3) exports NFATc proteins from the nucleus^{8,10,11} (see overview in Supplementary Fig. 1).

Phenotypes of *Nfatc*-null mice

Specific facial features are characteristic of Down's syndrome and arise from changes in embryonic bone development^{6,12,13}. *Nfatc2*^{−/−}; *Nfatc4*^{−/−} double-knockout mice have significantly reduced length

between the intersection of the parietal and interparietal bones and the nasale, a narrowed gap between the anterior aspects of the zygomatic arches, and shortened anterior parts of the skull (Fig. 1a–c and Supplementary Figs 2, 3). These animals display significantly shortened distances between the inferior-most point on the alveolar rim at the bone–tooth junction and both the mandibular angle and the posterior-most point on the mandibular condyle, but distances between posterior mandibular landmarks are less shortened (Supplementary Fig. 4). These characteristics of *Nfatc2*^{−/−}; *Nfatc4*^{−/−} mice resemble those observed in human Down's syndrome.

Individuals with Down's syndrome have cognitive deficits and muscular hypotonia, and often have sociable personalities^{6,12,14}. Previous work has shown that mice with forebrain-specific deletion of the protein phosphatase calcineurin B1 (*Cnb1*, also known as *Ppp3r1*), the activity of which is regulated by DSCR1, have defects in learning and memory¹⁵. In addition, calcineurin/NFAT signalling is essential for axonal outgrowth in response to neurotrophins and netrins¹⁶ during embryogenesis, and NFATc4 is a survival factor for cerebellar granule cells¹⁷, a cell population that is decreased in Down's syndrome individuals¹² and mouse models¹⁸. NFAT signalling also has an established role in myogenesis^{8,19}, and we found that certain interneuron subpopulations fail to develop in *Nfatc2*^{−/−}; *Nfatc3*^{−/−}; *Nfatc4*^{−/−} triple-knockout mice (H.W., I.A.G. and G.R.C., submitted manuscript). These findings suggest that even minor impairments in NFAT signalling might be sufficient to produce cognitive, behavioural and neuromuscular defects. In addition, we find that *Nfatc2*^{−/−}; *Nfatc4*^{−/−} mice show increased social interaction, increased locomotor activity, decreased muscular strength and decreased anxiety-related behaviour relative to control mice (Fig. 1d–g, Supplementary Fig. 5g–i). These characteristics are similar to those observed in Down's syndrome^{12,14}.

¹Department of Pathology, ²Program in Immunology, ³Division of Cardiovascular Medicine, Department of Medicine, ⁴Department of Developmental Biology, ⁵Department of Genetics and ⁶Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA. ⁷Genetic Engineering and Functional Genomics Unit, HMRO, Kyoto University Graduate School of Medicine, Kyoto 606-8501, Japan.

*These authors contributed equally to this work.

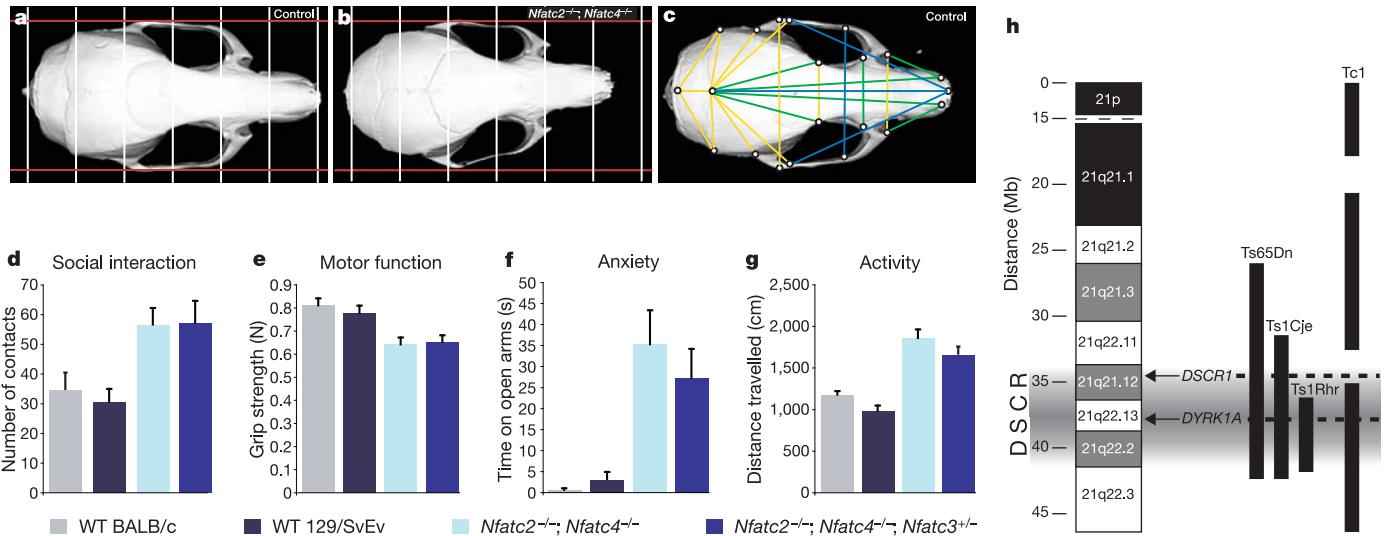


Figure 1 | Down's syndrome phenotypes in mice with mutations in the NFAT pathway. **a–c**, Superior view of crania from control (**a**, **c**) and *Nfatc2*^{−/−}; *Nfatc4*^{−/−} double-knockout (**b**) mice. **d**, NFATc mutant mice showed increased social interactions compared to BALB/c (*P* = 0.0036) and 129/SvEv (*P* = 0.0008) controls. **e**, Grip strength of NFATc mutants was significantly weaker (*P* < 0.0001 versus BALB/c, *P* = 0.0009 versus 129/SvEv). **f**, NFATc mutants entered significantly more into the open arms

of an elevated plus maze and spent more time in open arms (*P* < 0.0001 versus BALB/c, *P* < 0.0001 versus 129/SvEv). **g**, NFATc mutants also showed an increase in locomotor activity (*P* < 0.0001 versus BALB/c, *P* < 0.0001 versus 129/SvEv). Error bars in **d–g** indicate s.e.m. **h**, Map of human chromosome 21q. The DSCR is indicated in shaded area. Bars on the right denote the extent of the conserved orthologous region of MMU16 or transchromosomal HSA21 triplicated in Down's syndrome model mice.

Table 1 summarizes additional studies that we have conducted and compares features of Down's syndrome with the phenotypes of NFATc mutants and mouse models of Down's syndrome. In the present study, we find that NFATc mutant mice manifest the following characteristic features of Down's syndrome: placental vascular abnormalities (leading to the death of most fetuses with Down's syndrome; Supplementary Fig. 5a, b), increased sociability (Fig. 1d and Supplementary Fig. 5g), hypotonia (Fig. 1e and Supplementary Fig. 5h), an annular pancreas (Supplementary Fig. 5c, d) and an aganglionic megacolon (Supplementary Fig. 5e, f). We and others have shown that NFATc mutant mice manifest vascular and cardiac morphogenic defects^{20,21}, delayed tooth eruption (M.M.W. and G.R.C., submitted manuscript), behavioural changes (Fig. 1), a tendency to diabetes (J.J.H. and S.K.K., submitted manuscript),

obstructive nephropathy²², muscular weakness^{8,19} and immuno-deficiencies^{7,8,23}. Segmental trisomy 16 mice with genomic triplications orthologous to parts of human chromosome 21 (HSA21) have been described. Certain features of Down's syndrome such as placental insufficiency, cardiac defects, genitourinary abnormalities and gastrointestinal malformations are not observed in segmental trisomic mice, but are present in NFATc mutant mice. Although no individual NFATc mutant mouse reproduces all Down's syndrome pathologies, the features of Down's syndrome seem to be mild forms of NFATc mutant phenotypes.

The Down's syndrome critical region

These observations led us to examine the human Down's syndrome critical region (DSCR; see Supplementary Discussion A) for genes

Table 1 | Summary of Down's syndrome-like phenotypes in NFATc mutant and partial trisomy 16 mice

Phenotype	Down's syndrome individuals	NFATc mutant mice	Ts65Dn ²⁶	Ts1Cje ²⁷	Ts1Rhr ³⁰	Tc1 ²⁹
Placenta	Fetal loss (31–54%)	<i>Cnb1</i> ^{−/−} <i>Nfatc3</i> ^{−/−} ; <i>Nfatc4</i> ^{−/−20*}	No	No	No	No
Cardiovascular	Endocardial cushion defects (65%) ^{12,14}	<i>Cnb1</i> ^{−/−} <i>Nfatc1</i> ^{−/−} <i>Nfatc3</i> ^{−/−} ; <i>Nfatc4</i> ^{−/−} <i>Nfatc2</i> ^{−/−} ; <i>Nfatc3</i> ^{−/−} ; <i>Nfatc4</i> ^{−/−} CsA treatment ^{20,21}	No	No	No	Minimal
Neurological	Cognitive defects, hypotonia, increased sociability ¹²	<i>Cnb1</i> ^{−/−} <i>Nfatc2</i> ^{−/−} ; <i>Nfatc3</i> ^{−/−} ; <i>Nfatc4</i> ^{−/−} <i>Nfatc2</i> ^{−/−} ; <i>Nfatc4</i> ^{−/−†} * CsA treatment ^{15,16}	Yes ⁶	Yes ⁶	NR	Yes
Gastrointestinal	11% (1.4% Hirschsprung disease, 1.4% annular pancreas) ^{12,14}	<i>Nfatc3</i> ^{−/−} ; <i>Nfatc4</i> ^{−/−*}	No	No	NR	NR
Skeletal	Brachycephaly, midface hypoplasia, delayed tooth eruption ¹²	<i>Nfatc2</i> ^{−/−} ; <i>Nfatc4</i> ^{−/−†} <i>Nfatc1</i> ^{−/−‡}	Yes ¹³ NR	Yes ²⁸ NR	No NR	Minimal NR
Immune	Decreased interleukin-2 production and T-cell proliferation ¹²	<i>Nfatc1</i> ^{−/−} <i>Nfatc3</i> ^{−/−} <i>Nfatc1</i> ^{−/−} ; <i>Nfatc2</i> ^{−/−} CsA treatment ^{7,8,23}	NR	NR	NR	Minimal
Genitourinary	Obstructive nephropathy (18%) ^{12,14}	<i>Cnb1</i> ^{−/−22}	NR	NR	NR	NR

CsA, cyclosporin A; NR, not recorded.
* See Supplementary Fig. 5.
† See Fig. 1 and Supplementary Figs 2–4.
‡ M.M.W. and G.R.C., submitted manuscript.

that might inhibit NFATc function. *DSCR1* is located at the centromeric border of the DSCR (Fig. 1h) and encodes an inhibitor of calcineurin/NFAT signalling^{24,25}. *Dscr1* is triplicated in Ts65Dn and Ts1Cje mice, which have Down's syndrome-like craniofacial defects, but not in Ts1Rhr mice or Tc1 mice, which lack such craniofacial defects^{12,13,26–30} (see Fig. 1h for an overview of the trisomic regions in Tn65Dn, Ts1Cje, Ts1Rhr and Tc1 mice). *DSCR1* is also expressed at higher levels in Down's syndrome fetuses²⁵.

We examined the 25–30 genes in the DSCR for other potential NFAT regulators and identified *DYRK1A*, which encodes a nuclear serine/threonine kinase³¹ (Figs 1h and 2) that primes substrates for

phosphorylation by GSK3. GSK3 phosphorylates NFATc proteins in the nucleus, resulting in their inactivation and export^{10,11}. *DYRK1A* is expressed at elevated levels in some human Down's syndrome fetal tissues. *Dyrk1a*-deficient mice have defects in central nervous system (CNS) development³², and overexpression produces neurodevelopmental defects³³. *Dyrk1a* is sensitive to gene dosage, as heterozygous mutant mice show changes in CNS development³².

We found that *DYRK1A* regulates calcineurin/NFAT signalling in response to fibroblast growth factor-8 (FGF8), which has critical roles in development (Fig. 2a). *DYRK1A*, but not ERK1, inhibits FGF8-mediated induction of NFAT activity. Moreover, *DYRK1A* synergizes with *DSCR1* to block NFAT-dependent transcription (Fig. 2a). We also investigated the role of *DYRK1A* in embryonic neurons responding to spontaneous Ca^{2+} channel activity, which is required for neural development^{34,35}. Spontaneous activity induces a 12-fold increase in NFAT activity, which is blocked by *DYRK1A* (Fig. 2a). The role of *DYRK1A* in antagonizing NFAT may be general, in that NFAT-dependent transcription in HEK-293T cells is sensitive to *DYRK1A* and *DSCR1* (Supplementary Fig. 6).

GSK3 is required for transcriptional inactivation and export of NFATc proteins^{10,11}. It targets the first serine of a SPxxSP motif only if the second serine has been previously phosphorylated by a priming kinase. We found that *DYRK1A* synergizes with GSK3 to inhibit NFAT-dependent transcription in cortical neurons (Fig. 2b). Furthermore, *DYRK1A*, but not a kinase-inactive mutant (*DYRK1A*(KI)) can phosphorylate NFATc4 and prime it for subsequent phosphorylation by GSK3 (Fig. 2c). Endogenous *DYRK1A* immunoprecipitated from bFGF-stimulated H19-7 cells can prime NFATc4 for phosphorylation by GSK3 (Fig. 2d). Two conserved motifs in the amino termini of NFATc proteins, the serine-rich region and the serine/proline repeats, are the major sites of phosphorylation. Serine-to-alanine mutation of critical serines in these regions renders NFATc proteins independent of calcineurin activity and constitutively nuclear¹⁰. Using serine-to-alanine mutants of the serine-rich region and serine/proline repeats of NFATc4 together with mass spectrometric analysis (Supplementary Fig. 7), we found that *DYRK1A* phosphorylates the serine/proline repeats of NFATc4, consistent with a role as a priming kinase for GSK3³⁶.

NFATc proteins are rapidly exported from the nucleus, allowing discrimination between brief and sustained Ca^{2+} signals³⁷. We found that *DYRK1A*-transfected neurons show more than threefold faster rates of NFATc4 export than those transfected with an empty vector, and almost twofold faster export rates than those transfected with GSK3 (Fig. 2e, f). Thus *DYRK1A* directs nuclear export of NFATc4. *DYRK1A*, but not *DYRK1A*(KI), prevents nuclear occupancy by NFATc1 (Supplementary Fig. 6) in HEK-293T cells. Thus, we conclude that *DYRK1A* reduces NFAT transcriptional activity by direct phosphorylation of NFATc proteins, which leads to their nuclear export.

Consequences of increasing *Dscr1* and *Dyrk1a* dosage

To determine whether increasing the dosage of *Dscr1* and *Dyrk1a* can reproduce features of Down's syndrome, we generated nine lines of double transgenic mice overexpressing *Dyrk1a* and *Dscr1* during embryonic development, and studied those with low levels of *DYRK1A* and *DSCR1* protein expression. We monitored cardiac defects, which occur in half of Down's syndrome individuals¹² and all *Nfatc2*^{-/-}; *Nfatc3*^{-/-}; *Nfatc4*^{-/-} triple-knockout²¹ and *Nfatc1*^{-/-} embryos^{21,38,39}. Overexpression of *DYRK1A* (Fig. 3a, lane 2) at levels that are 2–3-fold above the endogenous protein is sufficient to induce vascular defects and block heart valve development (Fig. 3b). This modest overexpression of *DYRK1A* alone also decreases endogenous *DSCR1* and NFATc4 protein levels (Fig. 3a, lane 2), indicating that *DYRK1A* can inhibit expression of NFAT target genes, as predicted for an NFAT export kinase. Expression of both *DYRK1A* and *DSCR1* to 1.5–2-fold above endogenous levels (Fig. 3a, lane 3) leads to failure of heart valve elongation at embryonic day (E)13.5 (arrow in Fig. 3d),

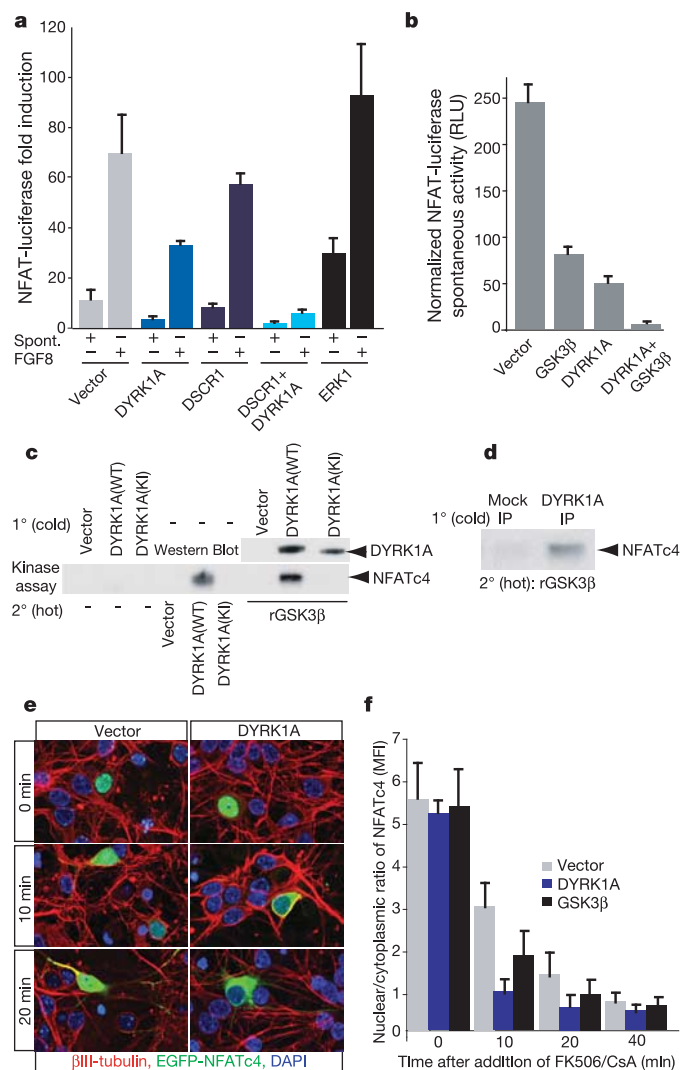


Figure 2 | *DYRK1A* is a nuclear export kinase for NFATc4. **a**, *DYRK1A* and *DSCR1* synergistically block NFAT-dependent transcription in cortical neurons. Expression of *DYRK1A* and/or *DSCR1* inhibits the activation of NFAT-dependent transcription in response to spontaneous activity (spont.) or stimulation by FGF8. **b**, *DYRK1A* and GSK3 β synergistically inhibit NFAT-dependent transcription in cortical neurons. **c**, Wild-type *DYRK1A*, but not a kinase-inactive mutant *DYRK1A*(KI), phosphorylates NFATc4 and primes it for phosphorylation by GSK3 β . 1° (cold), unlabelled ATP; 2° (hot), γ -³²P ATP. **d**, Endogenous *DYRK1A* immunoprecipitated from nuclear extracts can prime NFATc4 for GSK3 β phosphorylation. **e**, Accelerated nuclear export of an NFATc4-enhanced green fluorescent protein (EGFP) fusion protein after overexpression of *DYRK1A*, shown in representative confocal sections. Nuclear export was initiated by the addition of FK506/cyclosporin A (CsA) at $t = 0$ min. Original magnification, 40 $\times$. **f**, Quantification of NFATc4-EGFP cytoplasmic and nuclear mean fluorescence intensity (MFI) 0, 10, 20 and 40 min after the addition of FK506/CsA. Error bars indicate s.d.

comparable to that observed in *Nfatc1*^{-/-} embryos^{21,38,39}. Moreover, endocardially expressed NFATc1 is hyperphosphorylated and localized to the cytoplasm of these *Dyrk1a/Dscr1*-overexpressing animals (Fig. 3e–g).

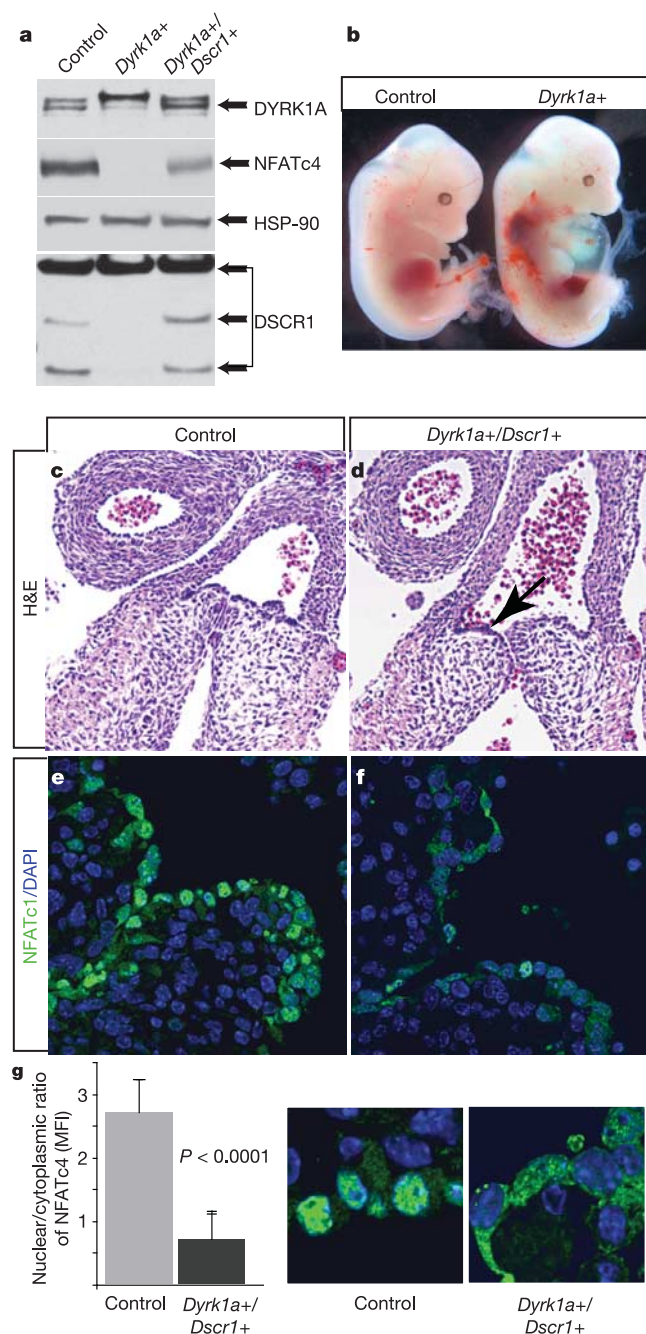


Figure 3 | Transgenic overexpression of *Dyrk1a* and *Dscr1* produces defects similar to NFAT mutants and Down's syndrome individuals.

a, Immunoblots of control (lane 1) and *Dyrk1a*-overexpressing (lane 2, *Dyrk1a*+) and *Dyrk1a/Dscr1*-overexpressing transgenic (Lane 3, *Dyrk1a*+/*Dscr1*+) E13.5 embryos. **b**, Gross morphology of an E13.5 control embryo (left) and a transgenic embryo transiently overexpressing *Dyrk1a* (right) with enlarged pericardial sac, vascular anomalies and heart failure. **c, d**, Blunted heart valves (arrow) in E13.5 *Dyrk1a/Dscr1*-overexpressing embryos. Original magnification, 20×. **e, f**, Confocal images showing redistribution of NFATc1 from the nucleus to the cytoplasm of endocardial cells in *Dyrk1a/Dscr1* double transgenic embryos. Original magnification, 40×. **g**, Quantification of the reduction in nuclear occupancy in *Dyrk1a/Dscr1* double transgenic embryos. Error bars indicate s.d. Immunofluorescent images are higher magnification insets of **e** and **f**.

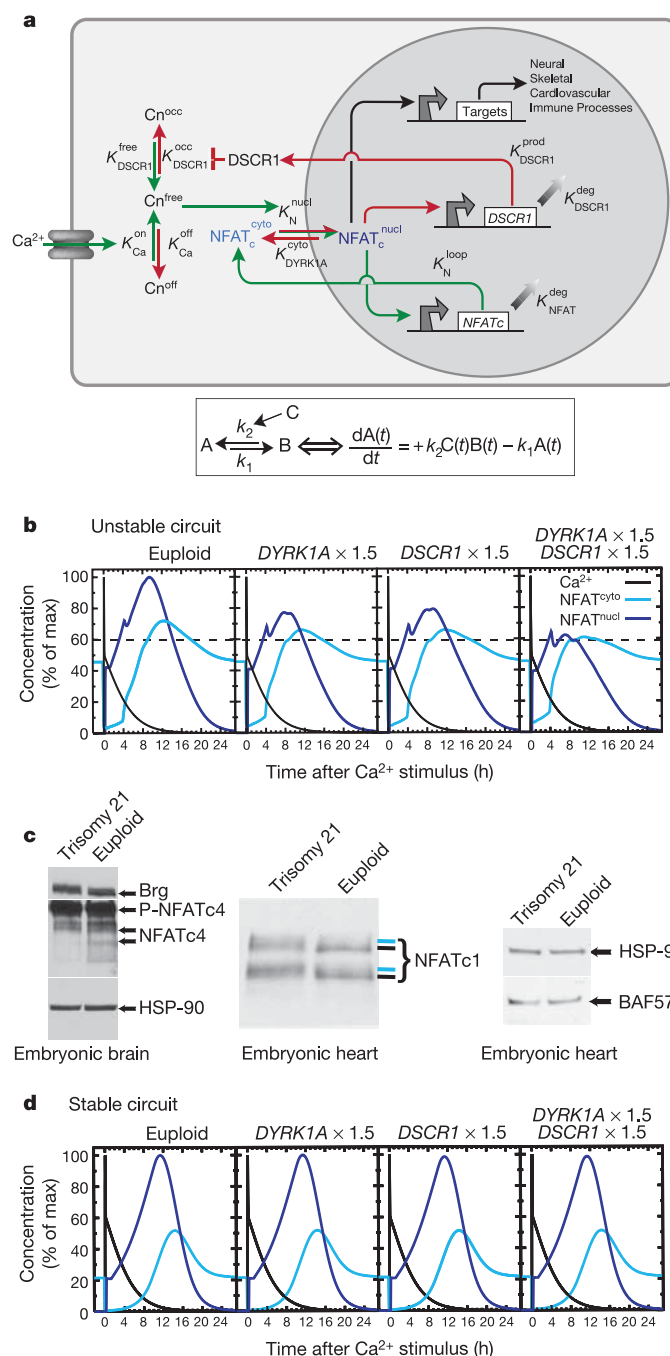


Figure 4 | Increased dosage of *DYRK1A* and *DSCR1* can significantly destabilize the NFAT regulatory circuit. **a**, The NFAT genetic regulatory circuit. Steps that increase the output of the genetic regulatory circuit are shown in green and steps that decrease the output of the circuit are in red. A generalized equation at the bottom is used as a template to calculate the change in concentration of each component over time. cyto, cytoplasmic (inactive) NFATc; deg, degradation; loop, positive or negative feedback loop; nucl, nuclear (active) NFATc; occ, occupied by DSCR1 (inactive enzyme); prod, production (gene transcription and translation). **b**, Simulation of NFAT regulatory circuit output under conditions of varying gene dosage of *DYRK1A* and *DSCR1* for an 'unstable circuit' in which nuclear NFATc levels are sensitive to a 1.5-fold increase in *DYRK1A* and *DSCR1* levels. Theoretical thresholds for gene activation are depicted at 60% of maximum. **c**, Hyperphosphorylated NFATc4 (P-NFATc4) and NFATc1 in 20-week trisomic human fetal brain and heart, relative to age-matched euploid tissues. **d**, Model of NFAT regulatory circuit output for a 'stable circuit' in which nuclear NFATc levels are relatively immune to 1.5-fold increases in *DYRK1A* and *DSCR1* levels.

We have found that NFATc4 positively regulates its own promoter through conserved NFAT binding sites in its regulatory region (Supplementary Fig. 8a, b). Positive feedback regulation is also consistent with the reduction in NFATc4 levels in *Dyrk1a/Dscr1*-overexpressing mice (Fig. 3a, lanes 2 and 3) and in *Cnbl*^{-/-} mice (Supplementary Fig. 8c). These results indicate that NFATc4, like NFATc1⁴⁰, regulates its own promoter (Supplementary Fig. 8d). These positive feedback loops should enhance the effects of a 1.5-fold increase in *DSCR1* and *DYRK1A* gene dosage in human Down's syndrome.

Mathematical modelling of the NFAT genetic regulatory circuit

These feedback buffering and amplification mechanisms, intrinsic to NFAT signalling, led us to use mathematical modelling to better predict the output of NFAT signalling (Fig. 4a). Because the initial level of each of the proteins involved in NFAT signalling is developmentally regulated in the tissues involved in Down's syndrome, we used experimentally determined ranges of starting conditions rather than a single level (Fig. 4a and Supplementary Discussion B). The initial Ca²⁺ signal was treated as a continuous profile to mimic spontaneous activity, which is a major regulator of both neural development^{34,35} and NFAT activity¹¹ (Fig. 2a). Two generalizations emerge. First, *DSCR1* and *DYRK1A* have additive effects under most initial concentrations of each component. Second, if we assume, as other studies have demonstrated⁴¹, that activating the promoters of target genes requires a threshold level of NFAT complexes, these genes may fail to be transcribed when *DSCR1* and/or *DYRK1A* are increased by 1.5-fold (Fig. 4b). Thus we predict that trisomy 21 will produce a disproportionate reduction in NFAT activity, resulting in a mild version of the aggregate defects observed in NFATc mutant mice.

To determine whether these predictions are borne out in mouse models of Down's syndrome, we examined two models with segmental trisomy^{26,27}. In cortical neurons of E13.5 Ts1Cje embryos, we found an increase in *DYRK1A* expression levels and also increased phosphorylation of NFATc4 (Supplementary Fig. 9). However, in whole heads of E11.5 Ts1Cje embryos and in hippocampal neurons of postnatal day (P)1 Ts65Dn mice we observed neither increased *DYRK1A* or *DSCR1* protein levels nor hyperphosphorylation of NFATc4 (Supplementary Fig. 9 and data not shown).

We also examined tissues from three human Down's syndrome fetuses and three age-matched controls at 17–21 weeks of gestation. During embryonic development, calcineurin/NFAT signalling is only required for brief and sharply defined time windows, and reduction in the activity of the pathway during these windows will phenocopy certain aspects of calcineurin/*Nfat*-null mutations^{16,20,21}. Although many NFAT-dependent developmental programmes have been executed by this stage, we still observed hyperphosphorylated forms of NFATc4 and NFATc1 in the brain and heart, respectively, of one 20-week trisomic fetus relative to control tissues (Fig. 4c). This finding is consistent with calcineurin inhibition by *DSCR1* and/or phosphorylation of NFATc by *DYRK1A*. Consistent with our model, which predicts that the NFAT genetic regulatory circuit may be unstable in some tissues (Fig. 4b, c) but stable in others (Fig. 4d), we found that NFATc1 phosphorylation was unchanged in spleen and muscle from the same fetus (not shown). These findings indicate that developmentally defined conditions are likely to lead to either stable or unstable states of the NFAT genetic regulatory circuit.

Discussion

Our studies suggest that under certain conditions, and perhaps for only brief developmental periods, increased dosage of *DSCR1* and *DYRK1A* in Down's syndrome reduces NFAT transcriptional activity, giving rise to mild versions of NFATc mutant phenotypes (Supplementary Fig. 1). Although certain post-developmental Down's syndrome pathologies, such as acute megakaryoblastic leukaemia and early-onset Alzheimer's disease, result from triplication

of other genes on HSA21 (*RUNX1* and *APP*, respectively), perturbation of the NFAT genetic regulatory circuit by increased dosage of *DSCR1* and *DYRK1A* may explain many of the developmental phenotypes in Down's syndrome. Here we were able to study only a limited number of trisomic tissues at a developmental stage that was perhaps past the critical period for NFAT function in development^{16,20,21}. Thus, additional work with human trisomic tissues from earlier developmental periods will be required to confirm our predictions.

Certain common but variable features of Down's Syndrome, such as heart disease, lethal placental vascular defects and immunodeficiency, are not seen in mouse models with segmental trisomy of murine chromosomal regions corresponding to HSA21^{12,26,27}. The absence of cardiovascular defects and immunodeficiency in the mouse models might reflect small differences in the kinetic constants between humans and mice shown in Fig. 4a. Subtle differences in these constants can be amplified in the positive and negative feedback loops, leading to substantial changes in the operation of the NFAT circuit. If this were the case, it would also explain the need to express *DSCR1* and *DYRK1A* at somewhat more than a 1.5-fold excess to give rise to heart defects in our studies. Mathematical simulations of the operation of the NFAT genetic regulatory circuit based on our current knowledge indicate that it is robust and will function well under the wide variety of initial concentrations that might be encountered in the development of different tissues and organs. However, it is particularly susceptible to an increase in the activity of the two synergistically functioning regulators of NFATc nuclear occupancy, *DSCR1* and *DYRK1A*, the genes for which happen to lie close to one another on HSA21. Our observations suggest that other human diseases may arise from the specific susceptibilities of genetic regulatory circuits, and that molecular understanding of these circuits will help to predict their weaknesses as well as possible sites of therapeutic intervention.

METHODS

Nfatc1-, *Nfatc2*-, *Nfatc3*- and *Nfatc4*-knockout mice have been described^{16,20,21}. Behavioural testing was performed as described⁴². E15.5 mouse cortical neurons were cultured and transfected as previously described¹⁶. Nuclear export assays of NFATc4 were performed as previously described¹¹. Rabbit polyclonal antibodies against *DYRK1A* were generated against recombinant rat *DYRK1A*. *In vitro* kinase assays were performed as previously described¹⁰. Transgenic embryos were generated by microinjection of full-length rat *Dyrk1a* and/or murine *Dscr1* cDNA under control of a β -actin promoter into fertilized oocytes.

The mathematical model was constructed by means of a coupled set of five first-order, nonlinear ordinary differential equations with gain and loss terms that were solved numerically with the standard fourth-order Runge–Kutta method, implemented with a C++ code of about 500 lines run on a Linux workstation.

Detailed methods can be found in the Supplementary Information.

Received 30 October 2005; accepted 27 February 2006.

Published online 22 March 2006.

- McAdams, H. H. & Shapiro, L. A bacterial cell-cycle regulatory network operating in time and space. *Science* **301**, 1874–1877 (2003).
- Davidson, E. H. *et al.* A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).
- Korenberg, J. R. *et al.* Down syndrome: toward a molecular definition of the phenotype. *Am. J. Med. Genet.* **7** (Suppl.), 91–97 (1990).
- Holtzman, D. M. & Epstein, C. J. The molecular genetics of Down syndrome. *Mol. Genet. Med.* **2**, 105–120 (1992).
- Korenberg, J. R. *et al.* Down syndrome phenotypes: the consequences of chromosomal imbalance. *Proc. Natl Acad. Sci. USA* **91**, 4997–5001 (1994).
- Antonarakis, S. E., Lyle, R., Dermitzakis, E. T., Raymond, A. & Deutsch, S. Chromosome 21 and Down syndrome: from genomics to pathophysiology. *Nature Rev. Genet.* **5**, 725–738 (2004).
- Graef, I. A., Chen, F. & Crabtree, G. R. NFAT signaling in vertebrate development. *Curr. Opin. Genet. Dev.* **11**, 505–512 (2001).
- Crabtree, G. R. & Olson, E. N. NFAT signaling: choreographing the social lives of cells. *Cell* **109** (Suppl.), S67–S79 (2002).
- Flanagan, W. M., Corthesy, B., Bram, R. J. & Crabtree, G. R. Nuclear association of a T-cell transcription factor blocked by FK-506 and cyclosporin A. *Nature* **352**, 803–807 (1991).

10. Beals, C. R., Sheridan, C. M., Turck, C. W., Gardner, P. & Crabtree, G. R. Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* **275**, 1930–1934 (1997).
11. Graef, I. A. *et al.* L-type calcium channels and GSK-3 regulate the activity of NF-ATc4 in hippocampal neurons. *Nature* **401**, 703–708 (1999).
12. Epstein, C. J. in *The Consequences of Chromosome Imbalance* (eds Barlow, P. W., Green, P. B. & Wylie, C. C.) (Cambridge Univ. Press, Cambridge, 1986).
13. Richtsmeier, J. T., Baxter, L. L. & Reeves, R. H. Parallels of craniofacial maldevelopment in Down syndrome and Ts65Dn mice. *Dev. Dyn.* **217**, 137–145 (2000).
14. Torfs, C. P. & Christianson, R. E. Anomalies in Down syndrome individuals in a large population-based registry. *Am. J. Med. Genet.* **77**, 431–438 (1998).
15. Zeng, H. *et al.* Forebrain-specific calcineurin knockout selectively impairs bidirectional synaptic plasticity and working/episodic-like memory. *Cell* **107**, 617–629 (2001).
16. Graef, I. A. *et al.* Neurotrophins and netrins require calcineurin/NFAT signaling to stimulate outgrowth of embryonic axons. *Cell* **113**, 657–670 (2003).
17. Benedito, A. B. *et al.* The transcription factor NFAT3 mediates neuronal survival. *J. Biol. Chem.* **280**, 2818–2825 (2005).
18. Baxter, L. L., Moran, T. H., Richtsmeier, J. T., Troncoso, J. & Reeves, R. H. Discovery and genetic localization of Down syndrome cerebellar phenotypes using the Ts65Dn mouse. *Hum. Mol. Genet.* **9**, 195–202 (2000).
19. Kegley, K. M., Gephart, J., Warren, G. L. & Pavlath, G. K. Altered primary myogenesis in *NFATC3*^{-/-} mice leads to decreased muscle size in the adult. *Dev. Biol.* **232**, 115–126 (2001).
20. Graef, I. A., Chen, F., Chen, L., Kuo, A. & Crabtree, G. R. Signals transduced by Ca²⁺/calcineurin and NFATc3/c4 pattern the developing vasculature. *Cell* **105**, 863–875 (2001).
21. Chang, C. P. *et al.* A field of myocardial-endocardial NFAT signaling underlies heart valve morphogenesis. *Cell* **118**, 649–663 (2004).
22. Chang, C. P. *et al.* Calcineurin is required in urinary tract mesenchyme for the development of the pyeloureteral peristaltic machinery. *J. Clin. Invest.* **113**, 1051–1058 (2004).
23. Hogan, P. G., Chen, L., Nardone, J. & Rao, A. Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev.* **17**, 2205–2232 (2003).
24. Rothermel, B. *et al.* A protein encoded within the Down syndrome critical region is enriched in striated muscles and inhibits calcineurin signaling. *J. Biol. Chem.* **275**, 8719–8725 (2000).
25. Fuentes, J. J. *et al.* DSCR1, overexpressed in Down syndrome, is an inhibitor of calcineurin-mediated signaling pathways. *Hum. Mol. Genet.* **9**, 1681–1690 (2000).
26. Reeves, R. H. *et al.* A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nature Genet.* **11**, 177–184 (1995).
27. Sago, H. *et al.* Ts1Cje, a partial trisomy 16 mouse model for Down syndrome, exhibits learning and behavioural abnormalities. *Proc. Natl Acad. Sci. USA* **95**, 6256–6261 (1998).
28. Richtsmeier, J. T., Zumwalt, A., Carlson, E. J., Epstein, C. J. & Reeves, R. H. Craniofacial phenotypes in segmentally trisomic mouse models for Down syndrome. *Am. J. Med. Genet.* **107**, 317–324 (2002).
29. O'Doherty, A. *et al.* An aneuploid mouse strain carrying human chromosome 21 with down syndrome phenotypes. *Science* **309**, 2033–2037 (2005).
30. Olson, L. E., Richtsmeier, J. T., Leszl, J. & Reeves, R. H. A chromosome 21 critical region does not cause specific Down syndrome phenotypes. *Science* **306**, 687–690 (2004).
31. Kentrup, H. *et al.* Dyrk, a dual specificity protein kinase with unique structural features whose activity is dependent on tyrosine residues between subdomains VII and VIII. *J. Biol. Chem.* **271**, 3488–3495 (1996).
32. Fotaki, V. *et al.* Dyrk1a haploinsufficiency affects viability and causes developmental delay and abnormal brain morphology in mice. *Mol. Cell. Biol.* **22**, 6636–6647 (2002).
33. Altafaj, X. *et al.* Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (*minibrain*), a murine model of Down's syndrome. *Hum. Mol. Genet.* **10**, 1915–1923 (2001).
34. Garaschuk, O., Linn, J., Eilers, J. & Konnerth, A. Large-scale oscillatory calcium waves in the immature cortex. *Nature Neurosci.* **3**, 452–459 (2000).
35. Borodinsky, L. N. *et al.* Activity-dependent homeostatic specification of transmitter expression in embryonic neurons. *Nature* **429**, 523–530 (2004).
36. Woods, Y. L. *et al.* The kinase DYRK phosphorylates protein-synthesis initiation factor eIF2B ϵ at Ser539 and the microtubule-associated protein tau at Thr212: potential role for DYRK as a glycogen synthase kinase 3-priming kinase. *Biochem. J.* **355**, 609–615 (2001).
37. Timmerman, L. A., Clipstone, N. A., Ho, S. N., Northrop, J. P. & Crabtree, G. R. Rapid shuttling of NF-AT in discrimination of Ca²⁺ signals and immunosuppression. *Nature* **383**, 837–840 (1996).
38. Ranger, A. M. *et al.* The transcription factor NF-ATc is essential for cardiac valve formation. *Nature* **392**, 186–190 (1998).
39. de la Pompa, J. L. *et al.* Role of the NF-ATc transcription factor in morphogenesis of cardiac valves and septum. *Nature* **392**, 182–186 (1998).
40. Northrop, J. P. *et al.* NF-AT components define a family of transcription factors targeted in T-cell activation. *Nature* **369**, 497–502 (1994).
41. Fiering, S. *et al.* Single cell assay of a transcription factor reveals a threshold in transcription activated by signals emanating from the T-cell antigen receptor. *Genes Dev.* **4**, 1823–1834 (1990).
42. Miyakawa, T. *et al.* Conditional calcineurin knockout mice exhibit multiple abnormal behaviors related to schizophrenia. *Proc. Natl Acad. Sci. USA* **100**, 8987–8992 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A summary figure is also included.

Acknowledgements We thank S. L. Schreiber, W. Mobley and K. Tanda for discussion and comments on the manuscript; W. Becker for *Dyrk1a* expression constructs; E. Olson for the anti-DCSR1 (MCIP1) antibody; R. S. Williams for the *Dscr1* (MCIP1) expression construct; K. Stankunas, G. Krampitz and C. Shang for help with histology on *Dyrk1a* and *Dscr1* transgenic mice; E. Wang for mass spectrometric analysis; W. Mobley and K. Zhan for providing Ts65Dn mice; and F. Wang, members of the Crabtree laboratory, J. Lee, M. Dionne and S. Arron for discussions. We thank the Stanford Center for Innovation in *In Vivo* Imaging (NCI Small Animal Imaging Resource Program Grant), the Stanford Imaging Facility and the Stanford Proteomics and Integrated Research Facility. These studies were supported by the Howard Hughes Medical Institute and NIH grants to G.R.C., and by the Christopher Reeve Paralysis Foundation (I.A.G.). M.M.W. is supported by a Stanford Graduate Fellowship and an HHMI predoctoral fellowship, C.-P.C. by grants from AHA and the NIH, J.R.A. by a postdoctoral fellowship from the Berry Foundation, J.J.H. and S.K.K. by the ADA, H.W. by a Damon Runyon Cancer Research Foundation postdoctoral fellowship and a Muscular Dystrophy Association research development grant, and T.M. by KAKENHI from JSPS and MEXT and by a grant from JST BIRD.

Author Contributions The order of listing of the authors J.R.A., M.M.W., A.P. and I.A.G. does in no way reflect their relative contribution to this work. I.A.G. and G.R.C. are responsible for the original concept. I.A.G. generated the *Nfatc* mutant mice (Fig. 1, Supplementary Figs 2–5 and Table 1), J.R.N. the *Cnb1* mutant mice (Supplementary Fig. 8c), and H.W. and L.C. the *Dyrk1a/Dscr1* transgenic mice (Fig. 3). I.A.G., M.M.W., C.-P.C., X.G., J.R.N., J.J.H., S.K.K., N.Y. and T.M. analysed mutant mice (Figs 1, 3 and Supplementary Figs 2–5 and Table 1). M.M.W. performed the skull morphometry studies (Fig. 1a–c and Supplementary Figs 2–4) and helped with the analysis of Ts1Cje and Ts65Dn mice (Supplementary Fig. 9). I.A.G. performed the neuron signalling experiments (Fig. 2a, b, e, f and Supplementary Figs 7, 8b), biochemical analysis of human Down's syndrome samples (Fig. 4c), calcineurin mutant mice (Supplementary Fig. 8c), Ts1Cje and Ts65Dn mice (Supplementary Fig. 9) and *Dyrk1a/Dscr1*-overexpressing mice (Fig. 3a) as well as the *Nfatc4* promoter studies (Supplementary Fig. 8). A.P. generated and solved the mathematical model (Fig. 4a, b, d and Supplementary Discussion B). U.F. provided the clinical samples (Fig. 4c). J.R.A. conducted the *in vitro* kinase (Fig. 2c, d and Supplementary Fig. 7) and 293T (Supplementary Fig. 6) assays and DSCR1/DYRK1A quantifications (used in Supplementary Discussion B), made the anti-DYRK1A antiserum and helped H.W. and I.A.G. to genotype some of the *Dyrk1a/Dscr1*-overexpressing mice. G.R.C., I.A.G., J.A.A., M.M.W. and A.P. wrote the manuscript and I.A.G., M.M.W., A.P. and G.R.C. generated the figures.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.R.C. (crabtree@stanford.edu) or I.A.G. (igraef@stanford.edu).

The Cenozoic palaeoenvironment of the Arctic Ocean

Kathryn Moran¹, Jan Backman², Henk Brinkhuis³, Steven C. Clemens⁴, Thomas Cronin⁵, Gerald R. Dickens⁶, Frédérique Eynaud⁷, Jérôme Gattacceca⁸, Martin Jakobsson², Richard W. Jordan⁹, Michael Kaminski¹⁰, John King¹¹, Nalan Koc¹², Alexey Krylov¹³, Nahysa Martinez¹⁴, Jens Matthiessen¹⁵, David McInroy¹⁶, Theodore C. Moore¹⁷, Jonaotaro Onodera¹⁸, Matthew O'Regan¹¹, Heiko Pälike¹⁹, Brice Rea²⁰, Domenico Rio²¹, Tatsuhiko Sakamoto²², David C. Smith¹¹, Ruediger Stein¹⁵, Kristen St John²³, Itsuki Suto²⁴, Noritoshi Suzuki²⁵, Kozo Takahashi¹⁸, Mahito Watanabe²⁶, Masanobu Yamamoto²⁷, John Farrell¹¹, Martin Frank²⁸, Peter Kubik²⁹, Wilfried Jokat¹⁵ & Yngve Kristoffersen³⁰

The history of the Arctic Ocean during the Cenozoic era (0–65 million years ago) is largely unknown from direct evidence. Here we present a Cenozoic palaeoceanographic record constructed from >400 m of sediment core from a recent drilling expedition to the Lomonosov ridge in the Arctic Ocean. Our record shows a palaeoenvironmental transition from a warm 'greenhouse' world, during the late Palaeocene and early Eocene epochs, to a colder 'icehouse' world influenced by sea ice and icebergs from the middle Eocene epoch to the present. For the most recent ~14 Myr, we find sedimentation rates of 1–2 cm per thousand years, in stark contrast to the substantially lower rates proposed in earlier studies; this record of the Neogene reveals cooling of the Arctic that was synchronous with the expansion of Greenland ice (~3.2 Myr ago) and East Antarctic ice (~14 Myr ago). We find evidence for the first occurrence of ice-rafted debris in the middle Eocene epoch (~45 Myr ago), some 35 Myr earlier than previously thought; fresh surface waters were present at ~49 Myr ago, before the onset of ice-rafted debris. Also, the temperatures of surface waters during the Palaeocene/Eocene thermal maximum (~55 Myr ago) appear to have been substantially warmer than previously estimated. The revised timing of the earliest Arctic cooling events coincides with those from Antarctica, supporting arguments for bipolar symmetry in climate change.

The significance of this palaeoenvironmental record is rooted in the Arctic Ocean's influence on global climate—specifically in terms of sea ice, which affects albedo, and the formation of cold, dense, bottom waters that drive global thermohaline circulation^{1,2}. Recent studies have demonstrated the Arctic Ocean's role in freshening the upper ~1.5 km of the northern North Atlantic Ocean by increased export of sea ice, increased freshwater supply from the Nordic seas, and a deepening of Arctic Intermediate Waters^{2,3}. So extracting a Cenozoic record that reveals the presence or absence of ice in the Arctic, its associated impact on the Earth's albedo, and the temporal variations of surface and deep ocean temperature and salinity, is of first-order importance. Here we present initial findings primarily focused on the occurrence and timing of Arctic Ocean ice in the form of the two primary cryosphere climate drivers, sea ice and icebergs—

indicative of large continental ice sheets rimming and extending into the basin. Further, we interpret the sea ice record, a proxy for Earth's albedo, to discern seasonal from perennial periods.

In 1961, Heezen and Ewing⁴ recognized that the Mid-Atlantic Ridge extends into the Arctic Ocean, where it becomes the Gakkel ridge. This led to the hypothesis that the Lomonosov ridge was a continental fragment that broke away from the Eurasian continental margin owing to spreading along the Gakkel ridge. Aeromagnetic surveys support this assumption, and suggest that initiation of Arctic seafloor spreading along the Gakkel ridge began during chron 24, ~57 Myr ago, the time of the Palaeocene/Eocene boundary^{5,6}. As rifting progressed, the Lomonosov ridge moved away from the continent, towards the pole, and subsided ~1,200 m, initially at a slow rate in the Palaeogene, followed by more rapid subsidence later

¹Graduate School of Oceanography & Department of Ocean Engineering, University of Rhode Island, Narragansett, Rhode Island 02882, USA. ²Department of Geology & Geochemistry, Stockholm University, Stockholm SE-106 91, Sweden. ³Laboratory of Palaeobotany & Palynology, Department of Biology, Utrecht University, 3584 DC Utrecht, The Netherlands. ⁴Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA. ⁵US Geological Survey, Reston, Virginia 20192, USA. ⁶Department of Earth Science, Rice University, Houston, Texas 77251, USA. ⁷Département de Géologie et Océanographie, Université Bordeaux, 33405 Talence, France. ⁸Department of Geophysics, CEREGE (CNRS), 13545 Aix-en-Provence, France. ⁹Department of Earth & Environmental Sciences, Yamagata University, Yamagata 990-8560, Japan. ¹⁰Department of Earth Sciences, University College London, London WC1E 6BT, UK. ¹¹Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882, USA. ¹²Norwegian Polar Institute, Polar Environmental Center, 9296 Tromsø, Norway. ¹³Department of Lithology & Geochemistry, VNIIOkeangeologia, St Petersburg 190121, Russia. ¹⁴Earth Sciences, Boston University, Boston, Massachusetts 02215, USA. ¹⁵Alfred Wegener Institute Foundation for Polar and Marine Research, D-27515 Bremerhaven, Germany. ¹⁶British Geological Survey, Murchison House, Edinburgh EH9 3LA, UK. ¹⁷Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA. ¹⁸Department of Earth and Planetary Sciences, Kyushu University, Fukuoka 812-8581, Japan. ¹⁹National Oceanography Centre, University of Southampton, Southampton SO14 3ZH, UK. ²⁰Department of Geography & Environment, University of Aberdeen, Aberdeen AB24 3UF, UK. ²¹Department of Geology, Paleontology & Geophysics, University of Padova, Padova 35137, Italy. ²²Institute for Research on Earth Evolution (IFREE), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka 237-0061, Japan. ²³Department of Geology & Environmental Science, James Madison University, Harrisonburg, Virginia 22807, USA. ²⁴Department of Earth & Planetary Sciences, Nagoya University, Chikusa, Nagoya 464-8602, Japan. ²⁵Institute of Geology & Paleontology, Tohoku University, Sendai City 980-8578, Japan. ²⁶Institute of Geoscience, Geological Survey of Japan, Ibaraki 305-8567, Japan. ²⁷Graduate School of Environmental Earth Science, Hokkaido University, Sapporo 060-0810, Japan. ²⁸Leibniz-Institute of Marine Sciences, IFM-GEOMAR, 24148 Kiel, Germany. ²⁹Paul Scherrer Institute, ETH-Hönggerberg, CH-8093 Zurich, Switzerland. ³⁰Department of Earth Science, University of Bergen, Allegaten 41, N-5007 Bergen, Norway.

in the Neogene. Hemipelagic and pelagic sediments accumulated atop the ridge during the Cenozoic. Seismic profiles across the ridge reveal a sediment sequence over 400 m thick on the ridge crest⁷, suggesting the presence of a unique archive of Cenozoic sedimentation and palaeoclimate history.

A Cenozoic geologic record from the Arctic Ocean

Our present understanding of this ocean basin is limited by the logistical difficulties of working in harsh, ice-covered regions, and is commensurate with our knowledge of the other ocean basins 50 years ago. Before the Integrated Ocean Drilling Program Arctic Coring Expedition⁸ (IODP ACEX), conducted in August 2004, the palaeoceanographic record of the central Arctic extended only to the mid-Pleistocene (~200–500 kyr ago), based on short piston cores, rarely longer than 10 m (ref. 9). Pre-Pleistocene material had rarely been recovered, and only piecemeal. The palaeoenvironmental record compiled from these short cores, although invaluable for reconstructions of relatively recent central Arctic glacial events, had not allowed the scientific community to address divergent hypotheses and a series of long-standing questions concerning the earlier Cenozoic evolution of the Arctic Ocean¹⁰. Closing this knowledge gap required a major technological advance, described elsewhere¹¹. In short, ACEX used two large icebreakers to enable a third, outfitted as a drill ship, to maintain position over a site in heavy, moving sea ice (>9/10 sea surface cover) for extended periods.

ACEX drilled and cored four sites on the Lomonosov ridge (Fig. 1). The sites were positioned along a site survey seismic reflection profile line⁷ (Fig. 2). This line was interpreted to represent a continuous Cenozoic sedimentary record atop rifted continental crust⁷. Although the sites are located up to ~15 km apart along the seismic line, the sediment cores from these sites could be correlated to each other on the basis of physical property data and on the continuity of seismic reflectors. This enabled us to create a composite sedimentary record that spans the Cenozoic. The recovered record comprises three distinct sediment types (in the upper ~200 m, 200–300 m and 300–420 m) and are described as separate units.

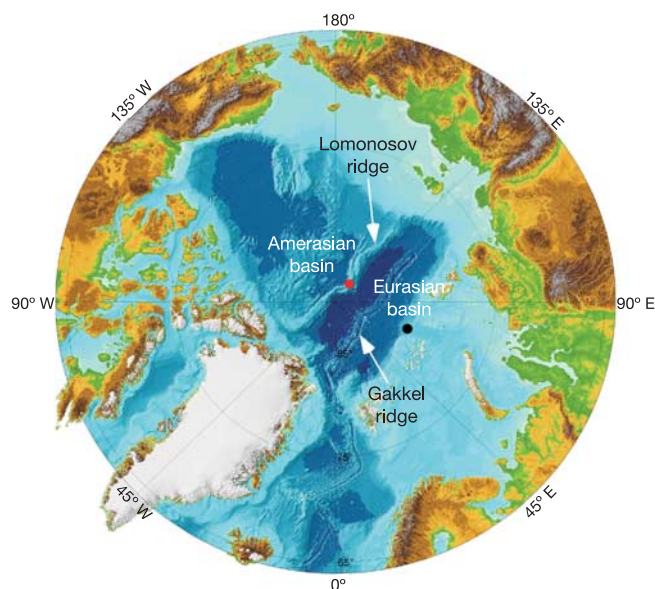


Figure 1 | Physiographic map of the Arctic Ocean. The location of the cores (red dot) is on the Lomonosov ridge, a continental sliver that broke away from the Eurasian continental margin ~57 Myr ago, owing to the spreading of the Gakkel ridge. Sediments atop the ridge record the palaeoenvironmental evolution of the Cenozoic Arctic Ocean. Colours are water depth; dark blue is ~4,000 m and the water depth over the ridge (light blue) varies from 800 m to 1,300 m. The black circle shows the approximate palaeoposition of the site. This map is from ref. 42.

The upper ~200 m (unit 1; Fig. 3a), which extends from the Miocene through to the present, consist of siliciclastic sediment, characterized by chemically controlled colour fluctuations, low organic carbon concentrations, and a biotic record dominated by organic walled microfossils (palynomorphs notably organic walled dinoflagellate cysts). Below 200 m, the early middle to early Eocene (44–50 Myr ago), sediments are microlaminated, organic carbon and biosiliceous rich, varying from silty clay to ooze (Fig. 3a; unit 2). Below 300 m, the early Eocene to late Palaeocene (50–56 Myr ago) sediment is predominantly siliciclastic and contains silica (only in the upper tens of metres) that has been altered to cristobalite. This interval (unit 3a; Fig. 3) unconformably overlies a Cretaceous unit composed of sand, sandstone and mudstone at about 400 m. The boundary is directly correlated to the seismically derived regional unconformity (Fig. 2) on the basis of conversion of the drilled depth to seismic time using core- and log-measured acoustic velocity.

The 'greenhouse' Arctic

The early Palaeogene marine environment can be characterized as warm, ice-free, brackish and biologically productive, resulting in the accumulation of siliciclastic sediments with high (up to 14%) organic carbon contents. The organic matter in the oldest part of the record (late Palaeocene) was mainly derived from algae, whereas that in the younger section (early middle Eocene) also includes higher plant material. This early middle Eocene section includes intervals of millimetre-scale laminated, pyritic black shale indicative of a quiet water anoxic basin¹², consistent with a shallow water environment.

In the oldest part of the sedimentary record, during the earliest Eocene, the typically warm water dinoflagellate genus *Apectodinium* dominated the fossil record in the pyrite-rich mudstone cores of unit 3. A global increase in *Apectodinium* occurred during the Palaeocene/Eocene thermal maximum (PETM)¹³, the largest known climatic warming of the Cenozoic. In a companion paper¹⁴, by TEX₈₆ analysis, we show that even at extreme high latitudes in the Arctic Ocean, peak PETM sea surface temperatures soared to ~24 °C. Additional studies of this interval show major changes in the hydrology of the region during the PETM warming¹⁵. Within and below the PETM interval, dinoflagellate cysts and agglutinated benthic foraminifera assemblages typify a shallow (neritic) marine depositional environment during the latest Palaeocene through to the early Eocene (Fig. 3).

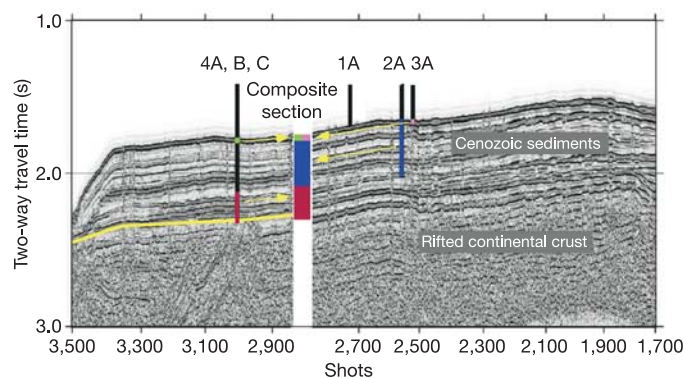


Figure 2 | Cross-section of the Lomonosov ridge. A seismic profile (AWI 91090) was interpreted as continental crust truncated by a regional unconformity overlain by a continuous sediment sequence⁷. The sites were positioned on this profile, shown as solid vertical lines. At each of the four sites (1–4), multiple holes (A, B, C...) were drilled and sampled to the depth of the solid vertical line. A 428-m-long composite sedimentary section was constructed by combining the holes, shown by different colours. The palaeoceanographic record overlies acoustic basement, a seismic unconformity (highlighted in yellow), that was confirmed to be Cretaceous continental crust.

A shift to an 'icehouse' Arctic

Later in the record, at ~ 49 Myr ago, a massive occurrence of glochidia (hair-like projections) and massulae of the fresh-water hydropterid fern *Azolla* confirms the presence of fresh surface water conditions¹⁶ with cooler TEX₈₆ temperatures of $\sim 10^\circ\text{C}$ (ref. 17). A concomitant occurrence of ebridian and silicoflagellate and diatom assemblages, requiring marine to brackish salinities, points to seasonally changing stratification regimes (Fig. 3). If the *Azolla* originated from a seasonal bloom, the associated fresh and relatively cool surface waters may have enabled winter ice formation, marking the start of a middle Eocene transition to a bipolar 'icehouse' world. The occurrence of a gneiss dropstone, 1 cm in diameter, at ~ 45 Myr ago is interpreted as ice-rafted debris (IRD), present much earlier than previously thought in the Northern Hemisphere. This dropstone was recovered from an undisturbed section of core, and could not have been reworked or moved downward from higher in the sedimentary section. The location of the ACEX site during this time period, although probably in shallow water (~ 200 m), was distal from the Siberian continental coast, and isolated from it by the Gakkel ridge, making improbable the delivery of pebbles to the site by any mechanism other than ice.

The sediment interval from ~ 44 to 16 Myr ago is characterized by slow to non-deposition and erosion that left a condensed and partially missing section. This interval marks a distinct change from organic-rich Palaeocene sediments with intervals of abundant microfossils, to

the sparsely fossiliferous silty muds of the overlying Neogene. The marked contrast represents a striking shift in the physical, chemical and ecological environment of the Arctic Ocean. This interval overlaps with the timing of a global shift from a largely ice-free, warm, 'greenhouse' world with high relative sea level to a world characterized by the climate-modulated waxing and waning of ice sheets.

The condensed interval ends in the early Miocene when sedimentation rates once again increased, and the unfossiliferous sediment was dominated by IRD, in the form of dropstones and the initial occurrence of sand (Fig. 3). The abundance of dropstones and sand suggests that sea ice and icebergs, calved from glaciers, were present in the Arctic Ocean. At ~ 14 Myr ago, the abundance of ice-derived sand increases significantly, with sedimentation rates reaching $1\text{--}2\text{ cm kyr}^{-1}$ and the intermittent appearance of cold- to cool-water dinoflagellates suggesting seasonal ice conditions.

At ~ 3.2 Myr ago, sediment physical properties show a large influx of coarse-grained sediment that is reflected in the consolidation index and acoustic velocity¹⁶ (Fig. 3c), suggesting increased sea ice and icebergs. This timing coincides with the start of large, Northern Hemisphere continental ice sheet expansion¹⁸, particularly in Greenland (Fig. 3b).

The youngest part of the ACEX record shows a change to slightly higher sedimentation rates at ~ 1 Myr ago (Fig. 3b), coincident with an intensification of the 100-kyr Milankovitch climate cycle. The

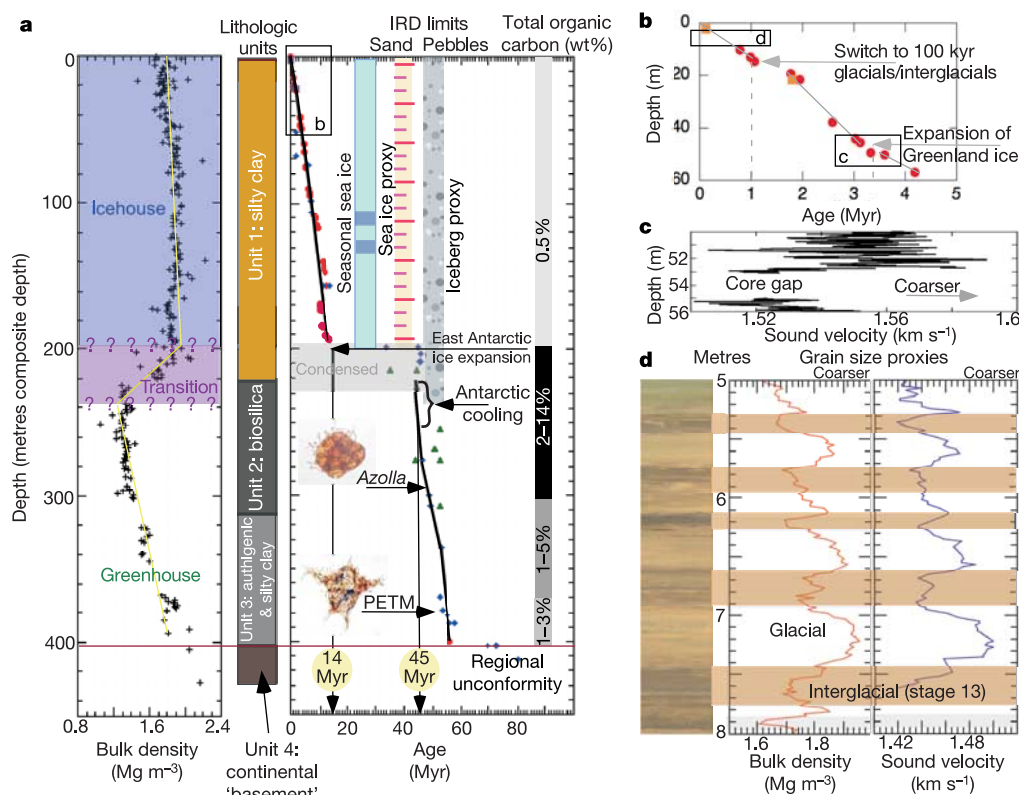


Figure 3 | Synthesis of the ACEX coring results. **a**, Right panel. Age is based on: palaeomagnetic stratigraphy shown as red circles; biostratigraphic data (dinocysts, blue diamonds; silicoflagellates, green triangles; and a few calcareous microfossils, squares, which only occur in the upper 25 m and can best be seen in **b**). Box shows position of panel **b**. Two micrographs are shown in their stratigraphic positions; the upper is *Azolla* and the lower is *Apectodinium*. Middle panel, lithologies are labelled with separate colours. Left panel, the Eocene 'greenhouse' world is separated from the Miocene 'icehouse' world by a transition period whose timing is not well constrained. Bulk density shows a normal increase in depth during both 'greenhouse' and 'icehouse' times; yellow solid lines are consolidation trends, suggesting uninterrupted sedimentation. Ice-rafted debris (IRD) in the form of pebbles begins in the 'greenhouse', and IRD (sand) starts at ~ 14 Myr ago (as noted

by the black, labelled arrow, bottom of right panel, consistent with the timing of the East Antarctic ice expansion). Interpreted seasonal ice occurrences are noted as darker blue bars within lighter blue shading. The arrow labelled 45 Myr, bottom of right panel, indicates concurrence with the Antarctic cooling time span shown by the bracket above. PETM, Palaeocene/Eocene thermal maximum. See text for information on *Azolla*. **b**, The age model from 0 to 5 Myr ago, showing changes in sedimentation rates concurrent with global climate events. Symbols denote different age control markers and are the same as in **a**. Boxes show positions of panels **c** and **d**. **c**, An example of increased sediment coarsening at 3.2 Myr ago. **d**, An expansion of the youngest part of the record, showing a core photograph (left) with proxies (right); glacials in the photograph are lighter, coarser and thicker in comparison to interglacials.

cause of this change remains unknown, but is presumed to be a nonlinear response to glacial ice expansion. ACEX results show a change in sediment composition across this major climate switch, and glacial and interglacial cycles are clearly resolvable (Fig. 3d). Further analyses of this interval may shed light on the contributing mechanisms for this increase in strength in the 100-kyr cycle.

Palaeoenvironmental implications

Origin of the Lomonosov ridge. ACEX confirmed the previously hypothesized continental origin of the Lomonosov ridge by recovering core material deeper than the Cenozoic sedimentary sequence across the regional seismic unconformity (Fig. 2). Below the late Palaeocene mudstone, ACEX recovered Upper Cretaceous (Campanian age; ~80 Myr) sands, sandstone and mudstone that are interpreted to have been of shallow marine origin because of the presence of agglutinated foraminifera. This palaeoenvironment is consistent with the theorized origin of the ridge as part of the shallow Barents continental margin. The confirmation of continental rifting also validates the ridge as a palaeoenvironment that captured a long sediment record over a time period of ~57 Myr.

Timing of palaeoenvironmental changes. ACEX results reveal changes during the PETM (~55 Myr ago). Our results show an early Cenozoic basin that was dominated by shallow water in the early Eocene and warm surface water conditions during the PETM¹⁴. These conditions have ramifications for explaining the causes of global warmth during this important time interval. A leading hypothesis for PETM warming is elevated atmospheric CO₂ (ref. 19) that probably briefly exacerbated the release of seabed methane hydrates, caused by a switch of deep water formation from high-latitude Southern Hemisphere regions to intermediate depth waters in the mid-latitudes of the North Pacific Ocean²⁰. ACEX results show that the Arctic Ocean experienced persistent surface water temperatures of 18 °C immediately before and after the PETM, but during this event temperatures peaked near 24 °C (ref. 14), which is notably higher than previous estimates of 10–15 °C (ref. 21) and indicates an even lower equator-to-pole temperature gradient than previously believed.

As important as extreme warmth is our evidence for a geographically isolated basin with a relatively wet climatic regime in the Arctic, as shown from palaeoplate reconstructions for this time²² and supported by the occurrence in ACEX cores of dark organic-rich sediment with sub-millimetre scale laminations, abundant fish teeth and bone fragments, the ecology of dinoflagellate and siliceous microfossil assemblages, and the lack of burrows and benthic microfauna. This evidence reveals a semi-enclosed basin environment with estuarine circulation and short-term (perhaps seasonal) oscillations in fresh to brackish conditions that reflect an absence of hyper-saline or evaporative conditions. At least at times, there was strong salinity stratification leading to seasonal or even perennial bottom-water hypoxia and anoxia. These indicators signify strong seasonality or inter-annual variability in runoff from adjacent continents, rather than forcing by tectonic processes or sea level oscillations. Two aspects of the globally warm PETM puzzle are clarified by these results and should be integrated into future models. First, wetter conditions and strong rainfall seasonality suggest an enhanced hydrological cycle in high latitudes of the Northern Hemisphere. Second, a PETM-isolated Arctic limits oceanic interchange between the Arctic and the North Atlantic²³ and decreases the likelihood that oceanic heat transport was directly responsible for warmth near the pole²⁴.

The increased occurrence of IRD at ~14 Myr ago in the ACEX cores suggests greater Arctic cooling and sea ice growth, which in turn, increases albedo and serves as a positive feedback for additional cooling. This timing is generally synchronous with the end of the middle Miocene climatic optimum and with East Antarctic ice sheet growth at 14.5 Myr ago²⁵, suggesting a second major, bipolar ice-growth event. The presence of Arctic Ocean sea ice may have

enhanced the albedo affect, contributing to low temperatures for subsequent Northern Hemisphere glacial ice expansion during the Pliocene. Mechanisms that had previously been proposed as responsible for glacial onset at 6–10 Myr ago (for example, enhanced run-off from Siberian rivers due to Tibetan uplift and an associated change in the freshwater budget of the Arctic's surface water)²⁶ may have contributed to glaciation, but post-date the initiation observed here.

Palaeoenvironmental evidence for sea ice and icebergs. Prior studies of Northern Hemisphere glaciation are based primarily on IRD in basins that are marginal to the Arctic Ocean, and thus can only indicate times when ice sheets were sufficiently large to shed icebergs into the neighbouring oceans and seas. The type of ice in the Arctic Ocean extends beyond icebergs to include sea ice, and possibly ice shelves. For example, recent sonar images show evidence of ice scouring of the sea floor on Lomonosov ridge, suggesting the presence of ice sheets up to 1 km thick²⁷ or similarly sized icebergs embedded in sea ice²⁸ in the central Arctic Ocean as recently as the early Pleistocene. ACEX results now provide direct evidence of icebergs (indicating the presence of proximal glaciers) and sea ice in the Arctic from the Eocene to the present.

A change from lower to higher organic carbon sediment is found at ~49 Myr ago, when the environment is interpreted to have been characterized by fresh, relatively cool (~10–14 °C) surface waters. Although a deep-water connection did not exist between the Arctic Ocean basin and the other oceans at this time, the presence of fresh water in the Arctic may have enhanced the initiation of sea ice that increased albedo and contributed to global cooling. This time interval is also characterized by a preservation of organic carbon in a shallow setting, in contrast to the modern Arctic Ocean continental shelves where organic carbon content deposition is relatively low, despite spring blooms.

In the Neogene record, the absence of dinoflagellate cysts (specifically intervals where the abundance is described as 'rare to absent') and the presence of IRD suggests perennial sea ice (Fig. 3a). Although the Arctic Ocean appears to have been ice covered during much of the Neogene, this was not always the case. Two intervals where dinoflagellate cyst abundance increases to 'common' occur at ~8 and ~9.2 Myr ago and could have been times of seasonal, rather than perennial, ice. This suggests that seasonal sea ice was present in the Arctic basin during the time when Greenland glacial ice began to grow. Continuing efforts to establish the geographic source of the IRD will allow verification, and improved differentiation between seasonal and perennial ice.

In the youngest portion of the record, during the late Quaternary the resolution is high enough to reveal late Pleistocene glacial-interglacial cycles (Fig. 3d). During glacial periods, larger grain size and concomitant increases in bulk density and sound velocity are attributable to intensified sea ice and iceberg activity.

Early onset of Northern Hemisphere ice. One striking feature of the overall sediment record is the average sedimentation rate ranging from 1 to 2 cm kyr⁻¹ (Fig. 3), which is an order of magnitude higher than estimates made by previous investigators who interpreted the Arctic as 'sediment-starved'^{29,29}. The higher sedimentation rates observed in the ACEX cores enable the reconstruction of palaeoceanographic records in sufficient detail to not only resolve glacial and interglacial cycles (Fig. 3d), but also to identify the timing of major climate events. The difficulty in developing a reliable chronostratigraphy formed the basis of a long-standing debate over sedimentation rates in this basin, with arguments made for both low (mm kyr⁻¹)^{29,30} and high (cm kyr⁻¹)^{9,31} rates. These disparate interpretations profoundly inhibited reconstructions of Arctic glacial history and are now resolved with the ACEX record, allowing for interpretations of the timing of northern high-latitude cooling events.

Investigators have long debated the timing, extent and nature of the onset of Northern Hemisphere glaciation^{32–38}. Previous studies suggest that such glaciation began between 10 and 6 Myr ago, whereas glaciation of Antarctica began much earlier, as early as ~43 Myr

ago³⁹. Our results push back the date of Northern Hemisphere cooling to ~45 Myr ago, and suggest contemporaneous co-evolution of ice on the poles, and thus symmetry in cooling and a bipolar transition from the 'greenhouse' to 'icehouse' world. Indirect evidence supporting the synchronicity of these events comes from detailed oxygen isotope records that document a sharp increase in $\delta^{18}\text{O}$ values across the Eocene/Oligocene boundary⁴⁰ and synchronous changes in calcium carbonate compensation depth⁴¹.

Arctic palaeoenvironmental significance

ACEX results suggest that the Earth's transition from the 'greenhouse' to the 'icehouse' world was bipolar, which points to greater control of global cooling linked to changes in greenhouse gases in contrast to tectonic forcing. Tectonic changes, such as the opening of the Drake Passage, may modify portions of the planet's ocean circulation systems but initiate climate changes that progress too slowly to induce synchronous global cooling patterns. Initial interpretations of the early presence of Arctic sea ice suggests that this mechanism may have led cooling in the Northern Hemisphere as early as ~45 Myr ago and is implicated as a participant in other cooling events, such as expansion of East Antarctic (~14.5 Myr ago) and Greenland (~3.2 Myr ago) ice. Further delineation of sea ice and its role in climatic cooling through enhanced albedo is warranted.

Received 4 August 2005; accepted 13 April 2006.

- Broecker, W. S. Thermohaline circulation, the Achilles heel of our climate system: will man-made CO₂ upset the current balance? *Science* **278**, 1582–1588 (1997).
- Holland, M. M., Bitz, C. M., Eby, M. & Weaver, A. J. The role of ice-ocean interactions in the variability of the North Atlantic thermohaline circulation. *J. Clim.* **14**, 656–675 (2001).
- Curry, R. & Mauritzen, C. Dilution of the northern North Atlantic Ocean in recent decades. *Science* **308**, 1772–1773 (2005).
- Heezen, B. C. & Ewing, M. in *Geology of the Arctic* (ed. Raasch, G.) 622–642 (Univ. Toronto Press, Toronto, 1961).
- Wilson, J. T. Hypothesis of the Earth's behaviour. *Nature* **198**, 925–929 (1963).
- Vogt, P. R., Taylor, P. T., Kovacs, L. C. & Johnson, G. L. Detailed aeromagnetic investigation of the Arctic basin. *J. Geophys. Res.* **84**, 1071–1089 (1979).
- Jokat, W., Uenzelmann-Neben, G., Kristoffersen, Y. & Rasmussen, T. ARCTIC'91: Lomonosov Ridge—a double sided continental margin. *Geology* **20**, 887–890 (1992).
- Backman, J., Moran, K., McInroy, D. & the IODP Exp. 302 Scientists, IODP Expedition 302, Arctic Coring Expedition (ACEX): A first look at the Cenozoic paleoceanography of the central Arctic Ocean. *Sci. Drilling*, **1**, 12–17 (2005).
- Backman, J., Jakobsson, M., Lovlie, R., Polyak, L. & Febo, L. A. Is the central Arctic Ocean a sediment starved basin? *Quat. Sci. Rev.* **23**, 1435–1454 (2004).
- Houghton, J. T. et al. (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, UK, 2001).
- Backman, J. Arctic Detailed Planning Group (ADPG) final report. *JOIDES J.* **27–2**, 18–27 (2001).
- Wignall, P. B. *Black Shales* (Oxford Univ. Press, Oxford, UK, 1994).
- Crouch, E. M. et al. Global dinoflagellate event associated with the late Palaeocene thermal maximum. *Geology* **29**, 315–318 (2001).
- Sluijs, A. et al. Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature* doi:10.1038/nature04668 (this issue).
- Pagani, M. et al. Arctic's hydrology during global warming at the Palaeocene/Eocene thermal maximum. *Nature* (submitted).
- Backman, J., Moran, K., McInroy, D. B., Mayer, L. A. & the Expedition Scientists. *Proc. IODP Exp. Rep. 302* (Integrated Ocean Drilling Program Management International, College Station, Texas, in the press).
- Brinkhuis, H. et al. Episodic fresh surface waters in the Eocene Arctic Ocean. *Nature* doi:10.1038/nature04692 (this issue).
- Shackleton, N. J. et al. Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. *Nature* **307**, 607–623 (1984).
- Sloan, L. C. & Barron, E. J. Eocene climate model results: Quantitative comparison to paleoclimatic evidence. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **93**, 183–202 (1992).
- Tripathi, A. & Elderfield, H. Deep-sea temperatures and circulation changes at the Paleocene-Eocene thermal maximum. *Science* **308**, 1894–1898 (2005).
- Tripathi, A., Zachos, J., Marinovich, L. Jr & Bice, K. Late Paleocene Arctic coastal climate inferred from molluscan stable and radiogenic isotope ratios. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **170**, 101–113 (2001).
- Lawver, L. A., Grantz, A. & Gahagan, L. M. in *Tectonic Evolution of the Bering Shelf-Chukchi Sea-Arctic Margin and Adjacent Landmasses* (eds Miller, E. L. A., Grantz, A. & Klemperer, S. L.) 333–358 (Special Paper 360, Geological Society of America, 2002).
- Knox, R. in *Late Paleocene-Early Eocene Climatic and Biotic Events in the Marine and Terrestrial Records* (eds Aubrey, M. P. et al.) 91–102 (Columbia Univ. Press, New York, 1998).
- Sloan, L. C. & Rea, D. K. Atmospheric CO₂ and Early Eocene climate: a general circulation study. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **119**, 275–292 (1995).
- Flower, B. P. & Kennett, J. P. Middle Miocene ocean-climate transition: high-resolution oxygen and carbon isotopic records from deep sea drilling project site 588A, southwest Pacific. *Paleoceanography* **8**, 811–844 (1993).
- Driscoll, N. W. & Haug, G. H. A short circuit in thermohaline circulation: a cause for Northern Hemisphere glaciation? *Science* **282**, 436–438 (1998).
- Polyak, L., Edwards, M. H., Coakley, B. J. & Jakobsson, M. Ice shelves in the Pleistocene Arctic Ocean inferred from glaciogenic deep-sea bedforms. *Nature* **410**, 453–456 (2001).
- Kristoffersen, Y. et al. Seabed erosion on the Lomonosov Ridge, central Arctic Ocean: A tale of deep draft icebergs in the Eurasia Basin and the influence of Atlantic water inflow on iceberg motion? *Paleoceanography* **19**, 3006–3019, doi:10.1029/2003PA000985 (2004).
- Clark, D. L., Whitman, R. R., Morgan, K. A. & Mackay, S. D. Stratigraphy and glacial marine sediments of the Amerasian basin, central Arctic Ocean. *Spec. Publ. Geol. Soc. Am.* **181**, 1–57 (1980).
- Witte, W. K. & Kent, D. V. Revised magnetostratigraphies confirm low sedimentation rates in Arctic Ocean cores. *Quat. Res.* **29**, 43–53 (1988).
- Nowaczyk, N. R. et al. Sedimentation rates in the Makarov Basin, central Arctic Ocean: a paleomagnetic and rock magnetic approach. *Paleoceanography* **16**, 368–389 (2001).
- Einarson, T., Hopkins, D. M. & Doell, R. R. in *The Bering Land Bridge* (ed. Hopkins, D. M.) 312–325 (Stanford Univ. Press, Stanford, 1967).
- Berggren, W. A., Kent, D. V., Swisher, C. C. III & Aubry, M.-P. in *Geochronology, Time Scales and Global Stratigraphic Correlation* (eds Berggren, W. A., Kent, D. V., Aubry, M.-P. & Hardenbol, J.) 129–212 (Spec. Publ., Soc. Sediment. Geol., Tulsa, 1995).
- Backman, J. Pliocene biostratigraphy of DSDP Sites 111 and 116 from the North Atlantic Ocean and the age of Northern Hemisphere glaciation. *Stockholm Contrib. Geol.* **32**, 115–137 (1979).
- Jansen, E., Fronval, T., Rack, F. & Channell, J. E. T. Pliocene-Pleistocene ice rafting history and cyclicity in the Nordic Seas during the last 3.5 Myr. *Paleoceanography* **15**, 709–721 (2000).
- Larsen, H. C. et al. *Proc. ODP Init. Rep. 152* (Ocean Drilling Program, College Station, Texas, 1994).
- Kleiven, H. F., Jansen, E., Fronval, T. & Smith, T. M. Intensification of Northern Hemisphere glaciations in the circum Atlantic region (3.5–2.4 Ma)—ice-rafted detritus evidence. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **184**, 213–223 (2002).
- Winkler, A., Wolf-Welling, T. C. W., Stattegger, K. & Thiede, J. Clay mineral sedimentation in high northern latitude deep-sea basins since the Middle Miocene (ODP Leg 151, NAAG). *Int. J. Earth Sci.* **91**, 133–148 (2002).
- Lear, C. H., Elderfield, P. A. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287**, 269–272 (2000).
- Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**, 53–57 (2005).
- Tripathi, A., Backman, J., Elderfield, H. & Ferretti, P. Eocene bipolar glaciation associated with global carbon cycle changes. *Nature* **436**, 341–346 (2005).
- Jakobsson, M., Cherkis, N., Woodward, J., Coakley, B. & Macnab, R. A new grid of Arctic bathymetry: A significant resource for scientists and mapmakers. *Eos* **81**(9), 89, 93, 96 (2000).

Acknowledgements We thank the IODP European Science Operator (ESO), led by the British Geological Survey; the Swedish Polar Research Secretariat, who provided shiptime for *Oden* as well as the ice and fleet management; icebreaker Captains Årnell, Backman, Davidjan, Haave, Shirley, Smith and Vikström, and the crews of *Oden*, *Vidar Viking* and *Sovetskiy Soyuz*; and Seacore Ltd, who provided the drilling services. This research used samples and data provided by the Integrated Ocean Drilling Program.

Author Contributions J.B. was the lead investigator on the ACEX proposal to IODP. K.M. and J.B. conducted much of the project planning with support from J.F. J.B. and K.M. led the expedition. Ship- and shore-based core analyses were conducted by the IODP 302 science party (H.B., S.C.C., T.C., G.R.D., F.E., J.G., M.J., R.W.J., M.K., J.K., N.K., A.K., N.M., J.M., T.C.M., D.M., J.O., M.O., H.P., B.R., D.R., T.S., D.C.S., R.S., K.S., I.S., N.S., K.T., M.W. and M.Y.). Additional sediment analyses were conducted by M.F. and P.K. Geophysical data were collected and analysed by W.J. and Y.K.

Author Information Reprints and permissions information is available at ngp.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.M. (kate.moran@gso.uri.edu).

LETTERS

Episodic fresh surface waters in the Eocene Arctic Ocean

Henk Brinkhuis¹, Stefan Schouten², Margaret E. Collinson³, Appy Sluijs¹, Jaap S. Sinninghe Damsté^{2,4}, Gerald R. Dickens⁵, Matthew Huber⁶, Thomas M. Cronin⁷, Jonaotaro Onodera⁸, Koza Takahashi⁸, Jonathan P. Bujak⁹, Ruediger Stein¹⁰, Johan van der Burgh¹, James S. Eldrett¹¹, Ian C. Harding¹², André F. Lotter¹, Francesca Sangiorgi^{1,2}, Han van Konijnenburg-van Cittert^{1,13}, Jan W. de Leeuw^{1,2}, Jens Matthiessen¹⁰, Jan Backman¹⁴, Kathryn Moran¹⁵ & the Expedition 302 Scientists†

It has been suggested, on the basis of modern hydrology and fully coupled palaeoclimate simulations, that the warm greenhouse conditions¹ that characterized the early Palaeogene period (55–45 Myr ago) probably induced an intensified hydrological cycle² with precipitation exceeding evaporation at high latitudes³. Little field evidence, however, has been available to constrain oceanic conditions in the Arctic during this period. Here we analyse Palaeogene sediments obtained during the Arctic Coring Expedition, showing that large quantities of the free-floating fern *Azolla* grew and reproduced in the Arctic Ocean by the onset of the middle Eocene epoch (~50 Myr ago). The *Azolla* and accompanying abundant freshwater organic and siliceous microfossils indicate an episodic freshening of Arctic surface waters during an ~800,000-year interval. The abundant remains of *Azolla* that characterize basal middle Eocene marine deposits of all Nordic seas^{4–7} probably represent transported assemblages resulting from freshwater spills from the Arctic Ocean that reached as far south as the North Sea⁸. The termination of the *Azolla* phase in the Arctic coincides with a local sea surface temperature rise from ~10 °C to 13 °C, pointing to simultaneous increases in salt and heat supply owing to the influx of waters from adjacent oceans. We suggest that onset and termination of the *Azolla* phase depended on the degree of oceanic exchange between Arctic Ocean and adjacent seas.

The recent Integrated Ocean Drilling Program (IODP) Expedition 302 (or Arctic Coring Expedition, ACEX) successfully recovered an expanded uppermost Palaeocene to middle Eocene sediment sequence from the Lomonosov ridge^{9,10} (Fig. 1). The sequence is barren of carbonate but comprises abundant organic matter, biosilica and pyrite⁹. Often, submillimetre laminations are present, while benthic microfossils are frequently absent⁹, evidencing sluggish deep water ventilation. Organic (mainly dinoflagellate cyst) and siliceous microfossil assemblages record an early Palaeogene Arctic Ocean with generally low surface water salinity. This trend culminates in a notable finding: the lowermost middle Eocene comprises microlaminated sediments with extraordinary abundances of microspore clusters

(massulae) of the free-floating, freshwater fern *Azolla* (Figs 2 and 3, and Supplementary Fig. S-2).

Nowadays, *Azolla* requires standing freshwater bodies, such as ponds, canals and flooded rice fields in tropical, subtropical, and warm temperate regions. Even when the most tolerant species have experimentally been pre-conditioned by gradual increase of salt concentrations, *Azolla* tolerates salinities only up to 5.5‰, but



Figure 1 | The early Eocene Arctic basin, site locations, and geographic distribution of the *Azolla* pulse in adjacent basins. Stars show pulse distribution. See Supplementary Fig. S-3 and Supplementary Table S-1 for more details.

¹Palaeoecology, Institute of Environmental Biology, Utrecht University, Laboratory of Palaeobotany and Palynology, Budapestlaan 4, 3584 CD Utrecht, The Netherlands. ²Royal Netherlands Institute for Sea Research (NIOZ), Department of Marine Biogeochemistry and Toxicology, PO Box 59, 1790 AB, Den Burg, Texel, The Netherlands. ³Department of Geology, Royal Holloway University of London, Egham, Surrey, TW20 0EX, UK. ⁴Department of Earth Sciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands. ⁵Department of Earth Sciences, Rice University, 6100 Main Street, Houston, Texas 77005, USA. ⁶Earth and Atmospheric Sciences Department and the Purdue Climate Change Research Center, Purdue University, 550 Stadium Mall Drive, West Lafayette, Indiana 47906, USA. ⁷USGS National Center, 12201 Sunrise Valley Drive, Reston, Virginia 20192, USA. ⁸Department of Earth and Planetary Sciences, Kyushu University, Fukuoka 812-8581, Japan. ⁹Bujak Research International, 105 North Park Drive, Blackpool, FY3 8NE, UK. ¹⁰Alfred Wegener Institute for Polar and Marine Research, Columbusstrasse, 27568 Bremerhaven, Germany. ¹¹Shell UK Exploration and Production, 1 Altens Farm Road, Nigg, Aberdeen, AB12 3FY, UK. ¹²School of Ocean & Earth Science, Southampton Oceanography Centre, University of Southampton, European Way, Southampton, SO14 3ZH, UK. ¹³National Museum of Natural History, 'Naturalis', PO Box 9517, 2300 RA Leiden, The Netherlands. ¹⁴Department of Geology and Geochemistry, Stockholm University, Stockholm, SE-10691, Sweden. ¹⁵University of Rhode Island, Bay Campus, Narragansett, Rhode Island 02882, USA.

†A list of authors and affiliations appears at the end of the paper.

normally *Azolla* cannot tolerate salinities higher than 1–1.6‰ (refs 11, 12). *Azolla* fossils are known from the mid-late Cretaceous onwards and occupied freshwater habitats, as reconstructed from a number of widely distributed complete fertile plants of several fossil species^{13,14}. The *Azolla*-rich sediments ($(5\text{--}30) \times 10^4$ remains per gram dry sediment) are characterized by a total organic carbon (TOC) content of ~5% (Fig. 2), compared to an early Palaeogene average of 3% (ref. 9). *Azolla* maxima co-vary with physical parameters such as sediment density and magnetic susceptibility and with TOC content, suggesting that in these intervals *Azolla* remains constitute a dominant organic component of the bulk sediment (Fig. 2, Supplementary Fig. S-2). At *Azolla* maxima, megaspores with or without attached microspore massulae and separate soral clusters of microspore massulae can be picked from sieved residues occurring at comparable abundances to those documented in Palaeogene microlaminated freshwater pond facies¹⁵ (Fig. 3, Supplementary Fig. S-2). On the basis of detailed light- and electron-microscopy studies of the *Azolla* microspore massulae, glochidia and megaspores, a morphological similarity is suggested with *Azolla areolata*, a species found only in Banks Island (District of Franklin) in the Eureka Sound group, Canadian High Arctic, in sediments with a tentative Palaeocene to Eocene age¹⁶. However, several distinctive features suggest that the specimens from the Lomonosov ridge probably constitute a new species of *Azolla*.

The palaeoecological and palaeoclimatological implications of the Arctic *Azolla* phase depend on whether the remains were brought in by periodic mass transport from freshwater bodies on adjacent continents, or if *Azolla* grew and reproduced *in situ* in the Arctic Ocean. The presence of mature megaspores with and without attached massulae, single, small groups and large clusters of massulae and probable aborted megaspores of *Azolla* in the ACEX material (Fig. 3, Supplementary Fig. S-2) all indicate that it is highly unlikely that this reflects a transported assemblage. This is strongly supported by the relative scarcity of terrestrially derived palynomorphs (pollen and spores). Further, values of the BIT (branched and isoprenoid tetraether) index, a measure for the amount of river-derived terres-

trial organic matter relative to marine organic matter¹⁷, are extremely low (<0.1), in marked contrast to values characteristic of the Palaeocene/Eocene thermal maximum that are found further down hole¹⁸. Collectively, this information reflects the distal position of the site and *in situ* *Azolla* growth. *Azolla* maxima in the bulk samples analysed from Cores 302-4A-11X and 302-4A-12X are accompanied by the lowest numbers and diversities of presumed exclusively marine plankton (Fig. 2) but high numbers of freshwater chrysophyte cysts (up to 1×10^7 per gram dry sediment; Fig. S-2). Hence, we conclude that surface waters of the central Arctic Ocean during the start of the middle Eocene were fresh enough to allow *in situ* growth and reproduction of *Azolla*.

However, brackish-marine dinocysts and notably diatoms, silicoflagellates and ebridians (up to 2×10^7 per gram dry sediment) continue to be present in all samples of the *Azolla* interval. Microscopy of the laminae indicates that the marine siliceous microfossils are concentrated in the light layers that are alternating with the dark, organic-rich *Azolla* layers (Supplementary Fig. S-2). Because the current age model⁹ allows the laminae to be possibly annual, they may reflect alternations of early spring brackish phytoplankton blooms followed by a late spring to summer precipitation increase, resulting in a stable stratified freshwater surface layer allowing rapid chrysophyte and *Azolla* growth. The frequency of the cyclic fluctuations in *Azolla* remains, density and magnetic susceptibility is about 1.3 m (Fig. 2). Given the average early Palaeogene accumulation rates of $\sim 1.3 \text{ cm kyr}^{-1}$ in Hole 4A (ref. 9), these oscillations may reflect orbital forcing of the $\sim 100\text{-kyr}$ eccentricity cycle, suggesting that the periodical freshening of the Arctic Ocean was astronomically modulated.

Abundant *Azolla* remains have previously been reported from the basal middle Eocene from all Nordic seas^{4–8}, predominantly through unpublished commercial oil and gas exploration studies (Supplementary Information). However, so far no particular attention has been given to the recognition, cause and implications of this unique phenomenon. The basal middle Eocene *Azolla* pulses in the Nordic seas are consistently associated with the last (abundant)

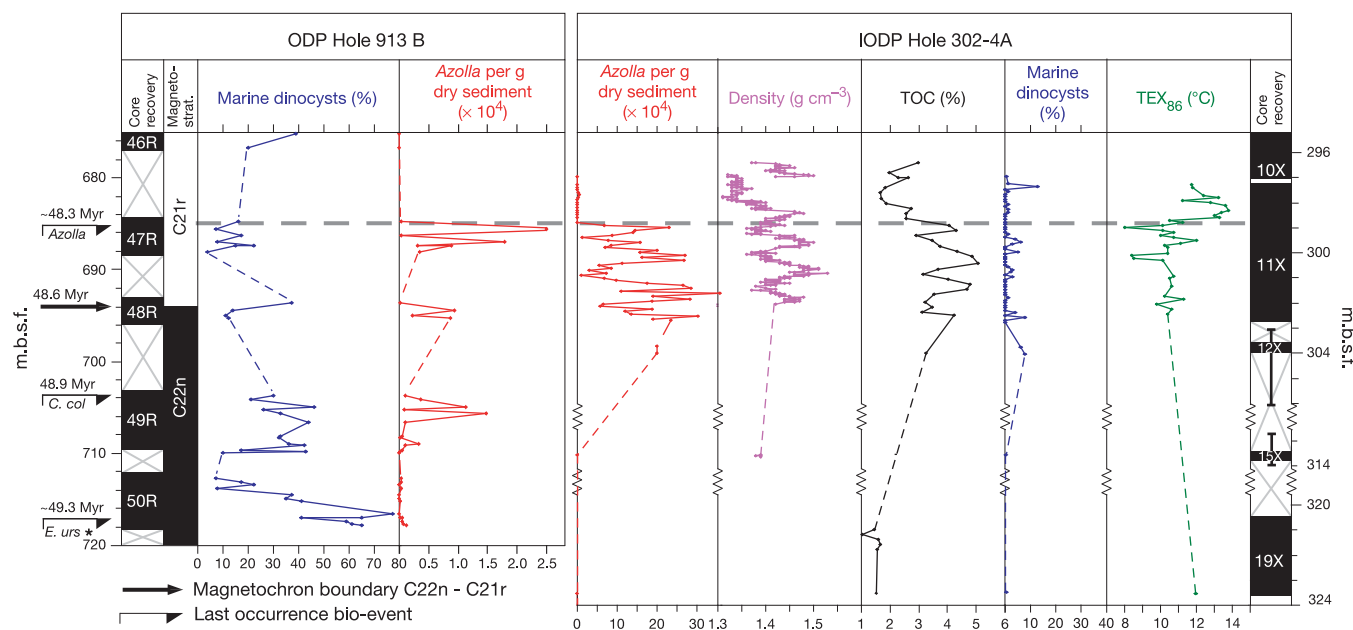


Figure 2 | Core recovery, chronology, palynological, physical properties and geochemical data across the *Azolla* phase of holes 302-4A and 913B. Correlation is based on the termination of the *Azolla* phase, and the general ACEX age model⁹. *E. urs.* and *C. col.* refer to the dinocysts *Eatonicystra ursulae* and *Charlesdownia columna*, respectively, and the asterisk indicates 'abundant'. TOC, total organic carbon; m.b.s.f., metres below sea floor;

TEX₈₆, an organic palaeothermometer. Numbers in core and recovery columns refer to core numbers. 'Magnetostat.' indicates magnetostratigraphy (that is, the measured direction of the palaeomagnetic field, and the number of the corresponding magnetochron recognized, Chron C22n or Chron C21r).

occurrences of two marine dinocyst species⁶ (Fig. 2), indicating that they are synchronous. The top of this interval (also referred to as the 'Megaspore sp. 1 Event', ref. 7), is widely used as a stratigraphic marker in industrial boreholes and outcrops^{5,6,8} (Fig. 1, Supplementary Fig S-3, Supplementary Table S-1). The termination of the *Azolla* phase has recently been shown to occur within magnetochron C21r (~48.3 Myr ago) at sites from the Greenland and Norwegian seas, notably in ODP Hole 913B (ref. 6; Fig. 2). The timing of its onset is less certain, but is estimated at around 49 Myr ago⁶. Although we have not yet been able to show involvement of identical *Azolla* species and the dinocyst markers are absent in the Arctic sediments, such an age assessment is well in line with the current age model for the Palaeogene ACEX cores, and we thus infer the pulses are time-correlative horizons across the Arctic and Nordic seas.

The number of *Azolla* remains per gram sediment in IODP Hole 4A (~25 × 10⁴) is ten times that in ODP Hole 913B (2.5 × 10⁴), while average sedimentation rates (~1.3 cm kyr⁻¹ at Hole 4A and ~3.0 cm kyr⁻¹ at Hole 913B; refs 6, 9) are only about 2.3 times lower in the Arctic. Hence, *Azolla* fluxes to the sea floor were significantly higher in the Arctic Ocean than in the adjacent seas. Also, at Hole 913B the *Azolla* interval is accompanied by much higher percentages of marine dinocysts than recorded on Lomonosov ridge (Fig. 2). In addition, the *Azolla* fluxes at Hole 913B and other Nordic sea sites (Supplementary Fig. S-3, Supplementary Table S-1) were still too

high to result from river input. These aspects support a scenario of Arctic basin *Azolla* mats being transported through huge Arctic freshwater plumes, to as far south as the southern North Sea basin⁸.

The onset as well as the demise of the *Azolla* phase in the Arctic Ocean potentially involved many factors that can change surface water conditions, such as temperature, salinity and mixing. We, therefore, estimated Arctic sea surface temperatures (SSTs) by applying the TEX₈₆ index, an organic palaeothermometer that is independent of salinity^{19–21} and calibrated to mean annual temperature. TEX₈₆ values suggest SSTs of ~10 °C during, and 13–14 °C immediately following, the *Azolla* phase (Fig. 2). These values are similar to, or slightly higher than, other late Palaeocene and Eocene floral, faunal and isotopic proxy evidence for mean annual temperature in the Arctic^{22–26}. Among several hypotheses to explain such high-latitude warmth, the leading (and not incompatible) two are increased heat transport and increased atmospheric greenhouse gas concentrations, as well as their associated feedbacks^{27–30}.

ACEX results for various time intervals indicate that relatively fresh conditions characterized the Arctic Ocean during the early Palaeogene^{9,10,18}, matching results from climate model experiments (Supplementary Fig. S-1), but only during an interval lasting ~800 kyr was the Arctic Ocean fresh enough to yield abundant *in situ* *Azolla* growth. To explain this pattern, we propose that the ~3 °C SST increase coincident with the termination of the *Azolla* phase represents a combined supply of heat and salt, resulting from the influx of waters from adjacent oceans. In contrast, the *Azolla* phase accompanied by relatively low temperatures (Fig. 2) suggests relatively limited combined supply of heat and salt. Such changes in heat and salinity are consistent in magnitude with the modern transports from the North Atlantic into the Arctic³, suggesting that these linked temperature–salinity signals reveal the turning on and off of a heat/salt transport switch. If correct, this implies that surface temperatures of ~11 °C (~20 °C warmer than today) prevailed in the Eocene Arctic without ocean heat transport. This indicates that increased greenhouse gas concentrations and associated feedbacks were probably the dominant factor in keeping high latitudes warm in the early Palaeogene.

METHODS

Palynology. Sediments were oven-dried at 60 °C. To ~2 g of sediment, a known amount of *Lycopodium* spores were added, after which the sample was treated with 30% HCl and twice with 30% HF for carbonate and silicate removal, respectively. After sieving over a 15-μm mesh sieve, residues were analysed at 500 × magnification. Absolute quantitative numbers were calculated using the relative number of *Lycopodium*.

Palaeobotany. A sediment sample (302-4A-11X, 111–113 cm) was disaggregated in hot water followed by short (5 min) treatment with 30% H₂O₂. The sample was then sieved through 125-μm mesh and *Azolla* megaspore and microspore massula clusters were picked from the particles retained on the sieve. Picked specimens were cleaned for 3 days in 60% HF, prepared for electron microscopy, and examined using a Hitachi H-7600 transmission electron microscope (TEM) and a FEI Quanta 200F field emission scanning electron microscope (SEM).

A portion of microlaminated sediment of the same level was embedded in epoxy resin and a polished thin section was prepared to standard thickness perpendicular to the laminations. The section was studied uncoated using an Hitachi S-3000N SEM in variable pressure mode at 70 Pa with a back scatter detector and elemental analysis was performed using an Oxford Instruments Link Isis series 300 energy dispersive X-ray (EDX) system. The same areas of the same section were also photographed using transmitted light microscopy and reflected plane polarized light microscopy.

Organic geochemistry. For the BIT and TEX₈₆ analysis, powdered and freeze-dried sediments (1–3 g dry mass) were extracted with dichloromethane (DCM)/methanol (2:1) by using the Dionex accelerated solvent extraction technique. The extracts were separated by Al₂O₃ column chromatography using hexane/DCM (9:1) and DCM/methanol (1:1) as subsequent eluents to yield the apolar and polar fractions, respectively. The polar fractions were analysed for tetraether lipids and BIT and TEX₈₆ (refs 17 and 19). TEX₈₆ values were converted to SST according to the equation: TEX₈₆ = (0.015 × SST) + 0.29 (ref. 19). Replicate analysis has shown that the error in TEX₈₆ values is ~0.01 or ~1 °C. TOC was determined using a LECO elemental analyser.

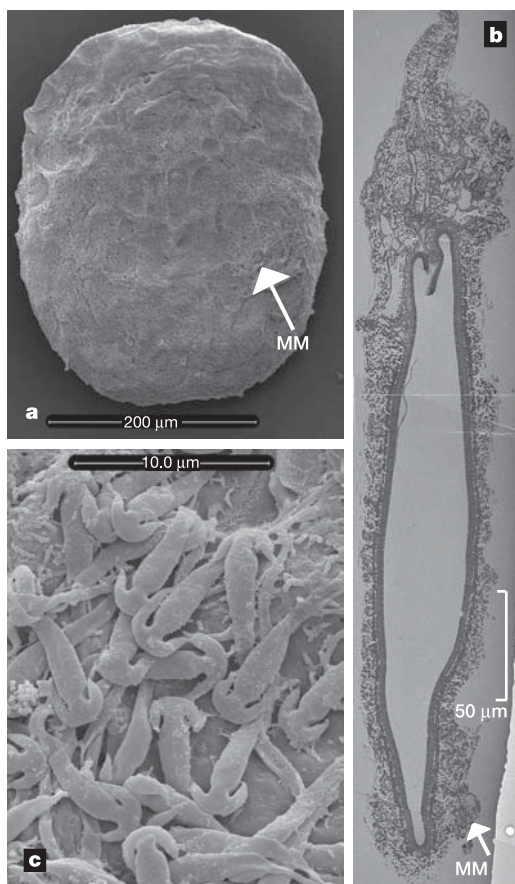


Figure 3 | *Azolla* mega- and microspores from Hole 302-4A-11X. **a**, SEM image of a megaspore apparatus, showing the outer dense alveolate wall and proximal cap of sporangial tissue (uppermost part of specimen); a microspore massula (MM) is attached. **b**, TEM image of a megaspore apparatus in longitudinal section, showing distal spore and proximal float zone, with fully developed spore wall and expanded spore; a microspore massula (MM) is attached. **c**, SEM image of fully developed glochidia on microspore massula showing anchor-shaped tips and distal dilation. See Supplementary Fig. S-2 for further documentation of *Azolla*.

Received 24 August 2005; accepted 23 February 2006.

- Zachos, J. C., Pagani, M., Sloan, L. C., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
- Barron, E. J., Hay, W. W. & Thompson, S. The hydrologic cycle: A major variable during Earth history. *Glob. Planet. Change* **1**, 157–174 (1989).
- Huber, M., Sloan, L. C. & Shellito, C. J. in *Causes and Consequences of Globally Warm Climates in the Early Palaeogene* (eds Wing, S. L., Gingerich, P. D., Schmitz, B. & Thomas, E.) 25–47 (GSA Special Paper 369, Geological Society of America, Boulder, Colorado, 2003).
- Boulter, M. C. Pollen and spore events from the marine Tertiary of North Europe. *J. Micropalaeontology* **5**, 75–84 (1986).
- Bujak, J. P. & Mudge, D. C. A high-resolution North Sea Eocene dinocyst zonation. *J. Geol. Soc. Lond.* **151**, 449–462 (1994).
- Eldrett, J. S., Harding, I. C., Firth, J. V. & Roberts, A. P. Magnetostratigraphic calibration of Eocene–Oligocene dinoflagellate cyst biostratigraphy from the Norwegian–Greenland Sea. *Mar. Geol.* **204**, 91–127 (2004).
- Gradstein, F. M. & Agterberg, F. in *Quantitative Stratigraphic Correlation* (eds Cubitt, J. M. & Reymont, R. A.) 119–175 (Wiley & Sons, Chichester, UK, 1982).
- Heilmann-Clausen, C. in *Early Paleogene Stage Boundaries. Abstracts and Field Trip Guides* (ed. Molina, E.) 19 (University of Zaragoza, Zaragoza, Spain, 1996).
- Backman, J., Moran, K., McInroy, D. B., Mayer, L. A. & the Expedition 302 Scientists. *Proc. IODP 302* doi:10.2204/iodp.proc.302.2006 (Integrated Ocean Drilling Program Management International, Inc., Edinburgh, 2006).
- Moran, K. *et al.* The cenozoic palaeoenvironment of the Arctic Ocean. *Nature* doi:10.1038/nature04800 (this issue).
- Rai, V. & Rai, A. K. Growth behaviour of *Azolla pinnata* at various salinity levels and induction of high salt tolerance. *Plant Soil* **206**, 79–84 (1998).
- Arora, A. & Singh, P. K. Comparison of biomass productivity and nitrogen fixing potential of *Azolla* spp. *Biomass Bioenergy* **24**, 175–178 (2003).
- Collinson, M. E. The ecology of Cainozoic ferns. *Rev. Palaeobot. Palynol.* **119**, 51–68 (2002).
- Collinson, M. E. Cainozoic ferns and their distribution. *Brittonia* **53**, 173–235 (2001).
- Collinson, M. E. Palaeofloristic assemblages and palaeoecology of the Lower Oligocene Bembridge Marls, Hamstead Ledge, Isle of Wight. *Bot. J. Linn. Soc.* **86**, 177–225 (1983).
- Sweet, A. R. & Hills, L. V. Early Tertiary species of *Azolla* subg. *Azolla* sect. *Krematospora* from western and arctic Canada. *Can. J. Bot.* **54**, 334–351 (1976).
- Hopmans, E. C. *et al.* A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. *Earth Planet. Sci. Lett.* **224**, 107–116 (2004).
- Sluijs, A. *et al.* Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum. *Nature* doi:10.1038/nature04668 (this issue).
- Schouten, S., Hopmans, E. C., Schefuß, E. & Sinninghe Damsté, J. S. Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth Planet. Sci. Lett.* **204**, 265–274 (2002).
- Wuchter, C., Schouten, S., Coolen, M. J. L. & Sinninghe Damsté, J. S. Temperature-dependent variation in the distribution of tetraether membrane lipids of marine Crenarchaeota: Implications for TEX₈₆ paleothermometry. *Paleoceanography* **19**, doi:10.1029/2004PA001041 (2004).
- Powers, L. A. *et al.* Crenarchaeotal membrane lipids in lake sediments: A new paleotemperature proxy for continental paleoclimate reconstruction? *Geology* **32**, 613–616 (2004).
- Greenwood, D. R. & Wing, S. L. Eocene continental climates and latitudinal temperature gradients. *Geology* **23**, 1044–1048 (1995).
- Markwick, P. J. Fossil crocodilians as indicators of Late Cretaceous and Cenozoic climates: Implications for using palaeontological data in reconstructing palaeoclimate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **137**, 205–271 (1998).
- Tripati, A., Zachos, J., Marincovich, L. Jr & Bice, K. Late Paleocene Arctic coastal climate inferred from molluscan stable and radiogenic isotope ratios. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **170**, 101–113 (2001).
- Jahren, A. H. & Sternberg, L. S. L. Humidity estimate for the middle Eocene Arctic rain forest. *Geology* **31**, 463–466 (2003).
- Fricke, H. C. & Wing, S. L. Oxygen isotope and paleobotanical estimates of temperature and $\delta^{18}\text{O}$ -latitude gradients over North America during the early Eocene. *Am. J. Sci.* **304**, 612–635 (2004).
- Huber, M. & Nof, D. The ocean circulation in the southern hemisphere and its climatic impacts in the Eocene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **231**, 9–28 (2006).
- Sloan, L. C. & Rea, D. K. Atmospheric carbon dioxide and early Eocene climate: A general circulation modeling sensitivity study. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **119**, 275–292 (1996).
- Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**, 600–603 (2005).
- Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406**, 695–699 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements H.B. thanks the Netherlands Organization for Scientific Research (NWO) and Utrecht University for enabling participation in the ACEx expedition. A.S. thanks the Utrecht Biogeology Centre for funding. M.H. thanks the Purdue Research Foundation for funding. This research used samples and data provided by the Integrated Ocean Drilling Program (IODP). We thank L. Bik, T. Brain, S. Gibbons, P. Goggin, N. Holloway, J. van Tongeren, N. Welters and M. Wolter for technical support and L.J. Lourens, H. Nohr-Hansen, M. Pagani, C.E. Stickley, G.L. Williams and J.C. Zachos for discussions.

Author Contributions H.B., A.S., F.S. and J.S.E. carried out the palynology, S.S. and J.S.D. the TEX₈₆ and BIT analyses, M.E.C., J.v.d.B. and H.v.K.-v.C. the palaeobotany, J.O. and K.T. generated siliceous microfossil numbers, while R.S. generated the TOC data. M.H. carried out the comparison to climate models. J.B. and K.M. were co-chiefs on the ACEx. J.P.B. and H.B. compiled the unpublished industrial palynological records. All authors contributed to writing the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.B. (H.Brinkhuis@bio.uu.nl).

†**Expedition 302 Scientists (those not listed above)** Steve Clemens¹, Frédérique Eynaud², Jérôme Gattacceca³, Martin Jakobsson⁴, Ric Jordan⁵, Michael Kaminski⁶, John King⁷, Nalân Koc⁸, Nahysa C. Martinez⁹, David McInroy¹⁰, Theodore C. Moore Jr¹¹, Matthew O'Regan⁷, Heiko Pälike¹², Brice Rea¹³, Domenico Rio¹⁴, Tatsuhiko Sakamoto¹⁵, David C. Smith⁷, Kristen E. K. St John¹⁶, Itsuki Suto¹⁷, Noritoshi Suzuki¹⁸, Mahito Watanabe¹⁹ & Masanobu Yamamoto²⁰

¹Geological Sciences, Brown University, 324 Brook Street, PO Box 1846, Providence, Rhode Island 02912-1846, USA. ²Department de Géologie et Océanographie, Université Bordeaux 1, Avenue des facultés, c/o Bernei Housen, 33405 Talence Cedex, France. ³Department of Geophysics, CEREGE (CNRS)/University of Aix-Marseille 3, BP 80, 13545 Aix-en-Provence Cedex 4, France. ⁴Department of Geology and Geochemistry, Stockholm University, 10691 Stockholm, Sweden. ⁵Department of Earth and Environmental Sciences, Faculty of Science, Yamagata University, 1-4-12 Kojirakawa-machi, Yamagata 990-8560, Japan. ⁶Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK. ⁷Graduate School of Oceanography, University of Rhode Island, Narragansett Bay Campus, Narragansett, Rhode Island 02882, USA. ⁸Norwegian Polar Institute, Polar Environmental Center, N-9296 Tromsø, Norway. ⁹Department of Earth Sciences, Boston University, 685 Commonwealth Avenue, Boston, Massachusetts 02215, USA. ¹⁰British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA, UK. ¹¹Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109-1063, USA. ¹²School of Ocean and Earth Science, University of Southampton, Southampton Oceanography Centre, European Way, Southampton SO14 3ZH, UK. ¹³Department of Geography and Environment, School of Geosciences, University of Aberdeen, Elphinstone Road, Aberdeen AB24 3UF, UK. ¹⁴Department of Geology, Paleontology & Geophysics, University of Padova, Via Giotto 1-1-35137 Padova, Italy. ¹⁵Institute for Research on Earth Evolution (IFREE), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Natsushima-cho 2-15, Yokosuka 237-0061, Japan. ¹⁶Department of Geology and Environmental Science, MSC 7703, James Madison University, Harrisonburg, Virginia 22807, USA. ¹⁷Institute of Life and Environmental Science, University of Tsukuba, Tennoudai 1-1-1, Tsukuba, Ibaraki 305-8572, Japan. ¹⁸Institute of Geology and Paleontology, Graduate School of Science, Tohoku University, Aramaki, Aoba, Aoba-ku, Sendai City 980-8578, Japan. ¹⁹Institute of Geoscience, National Institute of Advanced Industrial Science, and Technology (Geological Survey of Japan) AIST, Tsukuba Central 7, Higashi-1-1-1, Tsukuba, Ibaraki 305-8567, Japan. ²⁰Graduate School of Environmental Earth Science, Hokkaido University, Kita-10, Nishi-5, Kita-ku, Sapporo 060-0810, Japan.

LETTERS

Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum

Appy Sluijs^{1*}, Stefan Schouten^{2*}, Mark Pagani³, Martijn Woltering², Henk Brinkhuis¹, Jaap S. Sinninghe Damsté^{2,4}, Gerald R. Dickens⁵, Matthew Huber⁶, Gert-Jan Reichert⁴, Ruediger Stein⁷, Jens Matthiessen⁷, Lucas J. Lourens⁴, Nikolai Pedentchouk³, Jan Backman⁸, Kathryn Moran⁹ & the Expedition 302 Scientists†

The Palaeocene/Eocene thermal maximum, ~55 million years ago, was a brief period of widespread, extreme climatic warming^{1–3}, that was associated with massive atmospheric greenhouse gas input⁴. Although aspects of the resulting environmental changes are well documented at low latitudes, no data were available to quantify simultaneous changes in the Arctic region. Here we identify the Palaeocene/Eocene thermal maximum in a marine sedimentary sequence obtained during the Arctic Coring Expedition⁵. We show that sea surface temperatures near the North Pole increased from ~18°C to over 23°C during this event. Such warm values imply the absence of ice and thus exclude the influence of ice-albedo feedbacks on this Arctic warming. At the same time, sea level rose while anoxic and euxinic conditions developed in the ocean's bottom waters and photic zone, respectively. Increasing temperature and sea level match expectations based on palaeoclimate model simulations⁶, but the absolute polar temperatures that we derive before, during and after the event are more than 10°C warmer than those model-predicted. This suggests that higher-than-modern greenhouse gas concentrations must have operated in conjunction with other feedback mechanisms—perhaps polar stratospheric clouds⁷ or hurricane-induced ocean mixing⁸—to amplify early Palaeogene polar temperatures.

Stable carbon isotope ($\delta^{13}\text{C}$) records of carbonate and organic carbon from numerous sites show a prominent negative carbon isotope excursion across the Palaeocene/Eocene thermal maximum (PETM)^{2,9}. The carbon isotope excursion is expressed as a >2.5‰ drop in $\delta^{13}\text{C}$, which signifies an input of at least 1.5×10^{18} g of ^{13}C -depleted carbon, somewhat analogous in magnitude and composition to current and expected fossil fuel emissions. The PETM captures ~200 kyr (ref. 10) and is associated with profound environmental changes that are well-documented at low- to mid-latitudes (<60°), including a 4–8°C temperature rise of surface and deep ocean waters^{1–3} and major terrestrial and marine biotic changes^{11–13}. Terrestrial mammal turnovers are consistent with mass migrations across Arctic regions resulting from high-latitude warming¹⁴, but no Arctic data have existed to evaluate this hypothesis.

The Integrated Ocean Drilling Program Expedition (IODP) 302 (or the Arctic Coring Expedition) recently recovered a Palaeogene marine sedimentary record from Hole 4A (~87°52.00'N; 136°10.64'E; 1,288 m water depth), on the Lomonosov ridge in

the central Arctic Ocean⁵. This ridge represents a fragment of continental crust that rifted from the Eurasian shelf margin at high latitudes (>85°; Fig. 1) during the latest part of the Palaeocene epoch and subsided to present depths after the Palaeocene. Upper Palaeocene and Lower Eocene sediments between approximately 406 and 263 m composite depth below sea floor (m.c.d.) at Hole 4A consist of organic-rich (~2% total organic carbon (TOC) by mass on average) siliciclastic claystone⁵. Shipboard observations showed that this interval is barren of calcareous and siliceous microfossils but yields rich assemblages of palynomorphs, notably organic-walled dinoflagellate cysts (dinocysts) and terrestrial pollen and spores⁵.

The PETM was identified from the top of Core 32X to within Core

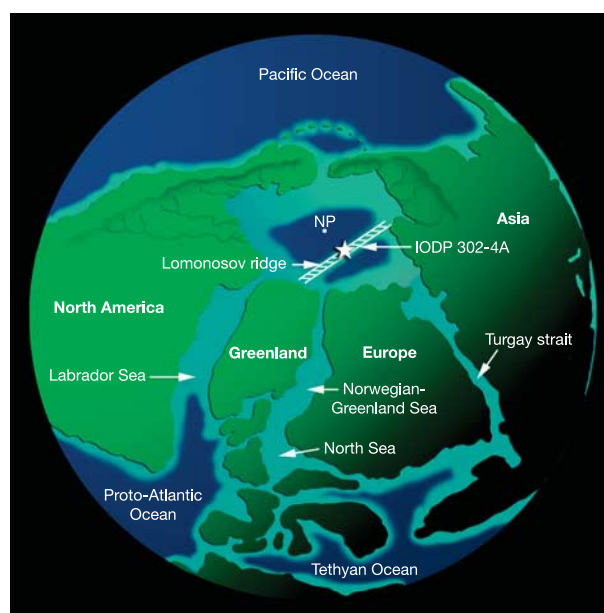


Figure 1 | Location of IODP Hole 302-4A within the palaeogeographic reconstruction of the Arctic Basin at late Palaeocene–early Eocene times. The figure is modified from ref. 24. NP, North Pole.

¹Palaeoecology, Institute of Environmental Biology, Utrecht University, Laboratory of Palaeobotany and Palynology, Budapestlaan 4, 3584 CD Utrecht, The Netherlands. ²Royal Netherlands Institute for Sea Research (NIOZ), Department of Marine Biogeochemistry and Toxicology, PO Box 59, 1790 AB, Den Burg, Texel, The Netherlands. ³Department of Geology and Geophysics, Yale University, PO Box 208109, New Haven, Connecticut 06520, USA. ⁴Department of Earth Sciences, Utrecht University, Budapestlaan 4, 3584 CD Utrecht, The Netherlands. ⁵Department of Earth Sciences, Rice University, 6100 Main Street, Houston, Texas 77005, USA. ⁶Earth and Atmospheric Sciences Department and the Purdue Climate Change Research Center, Purdue University, 550 Stadium Mall Drive, West Lafayette, Indiana 47906, USA. ⁷Alfred-Wegener-Institute for Polar and Marine Research, Columbusstrasse, 27568 Bremerhaven, Germany. ⁸Department of Geology and Geochemistry, Stockholm University, Stockholm, SE-106 91, Sweden. ⁹University of Rhode Island, Bay Campus, Narragansett, Rhode Island 02882, USA.

*These authors contributed equally to this work.

†A list of authors and affiliations appears at the end of the paper.

29X (~387–378.5 m.c.d.) by the occurrence of the dinocyst species *Apectodinium augustum*, which is diagnostic of the PETM¹⁵ (Fig. 2; Supplementary Fig. S-1a). The lower bound is somewhat problematic, though, because the upper 50 cm of Core 32X has been disturbed by drilling and various proxies suggests that the sediment from this interval represents a mixture of uppermost Palaeocene and PETM material¹⁶. Moreover, only 55 cm of section was recovered of the critical Core 31X, which has an uncertain stratigraphic position relative to Cores 30X and 32X (see error bars in Fig. 2 and Supplementary Information). Stable carbon isotopes of bulk organic carbon ($\delta^{13}\text{C}_{\text{TOC}}$) show a prominent ~6‰ drop between the top of Core 32X (388 m.c.d.) and 31X (~386 m.c.d.), apart from one value from the disturbed zone, followed by a gradual recovery through Cores 30X and 29X to ~378.5 m.c.d. (Fig. 2). The $\delta^{13}\text{C}_{\text{TOC}}$ pattern is generally reproduced in the carbon isotope record of the C_{27} and C_{29} *n*-alkanes, which are biomarkers derived from the leaf waxes of terrestrial higher plants¹⁶. Despite the core gaps, the magnitude and shape of the $\delta^{13}\text{C}_{\text{TOC}}$ excursion resembles other shallow marine PETM sections, such as Doel in Belgium¹⁷, confidently correlating this interval to the Palaeocene/Eocene boundary event.

Before the PETM, *Apectodinium* was a subtropical dinoflagellate restricted to low latitudes^{11,15}. Thus, the sudden influx of *Apectodinium* spp. dinocysts across the PETM at Hole 4A (Fig. 2) suggests a substantial rise in Arctic sea surface temperature (SST) to subtropical or tropical levels. Angiosperm pollen becomes more abundant at the expense of spores and gymnosperm pollen (Fig. 2), suggesting an expanded growing season. The lack of calcareous microfossils prohibits the use of the common techniques for quantifying past SSTs. Instead, we employed the newly developed palaeothermometer TEX_{86}' (see Supplementary Information), which is based on the distribution of crenarchaeotal membrane lipids¹⁸. This distribution is independent of surface water parameters such as nutrient availability or salinity^{18,19}, and shows a highly significant linear correlation with present-day mean annual SST in the range of 10 to 28 °C (Supplementary Fig. S-2B). Because the export of crenarchaeotal lipids to the sea floor predominantly occurs during the season with

highest phytoplankton productivity, which in the Arctic Ocean is summer, our TEX_{86}' record is probably skewed towards summer temperatures (see also Supplementary Information).

Arctic SSTs rose from ~18 °C in the latest Palaeocene, to over 23 °C during the PETM, and subsequently decreased to ~17 °C by the end of the event (Fig. 2). Background SSTs from the latest Palaeocene and early Eocene are generally consistent with the few other proxy data estimates from Arctic locations with late Cretaceous and early Palaeogene strata^{20–22}. The significantly lower terrestrial temperature estimates from Ellesmere Island at 73° N palaeolatitude²³ are derived from upper Lower Eocene strata and are similar to TEX_{86} -derived SSTs in the Arctic Ocean for that time period²⁴; these estimates are thus not in disagreement with our data. Maximum SSTs coincide with minimum $\delta^{13}\text{C}$ values during the PETM, while the cooling trend mirrors the recovery pattern in $\delta^{13}\text{C}$ and a decrease of angiosperm pollen.

Several lines of evidence (Fig. 2) suggest that the location of Hole 4A was proximal to the coast and strongly influenced by fluvial inputs in the latest Palaeocene. For example, palynomorph assemblages from upper Palaeocene strata are dominated by terrestrial spores and pollen (~90%). Those samples with sufficient dinocysts yield abundant *Senegalinium* spp. and *Cerodinium* spp. (Supplementary Fig. S-1), which probably come from dinoflagellates that tolerated low surface-water salinities²⁴ and required nutrient-rich conditions²⁵. Sediments from this interval also contain abundant amorphous organic matter, presumably of terrestrial origin. Moreover, values of the Branched and Isoprenoid Tetraether (BIT) index—a measure for the amount of river-derived terrestrial organic matter relative to marine organic matter²⁶—are relatively high.

In contrast to uppermost Palaeocene sediments, palynomorph assemblages from the PETM interval are characterized by abundant dinocysts (60%) and substantially lower BIT indices (Fig. 2), indicating a relative decrease of riverine-derived organic carbon. Also, the increase in the Rock Eval hydrogen index suggests a relative increase in aquatic versus terrestrial organic matter (Supplementary Fig. S-3A). We attribute the enhanced influence of marine conditions during the

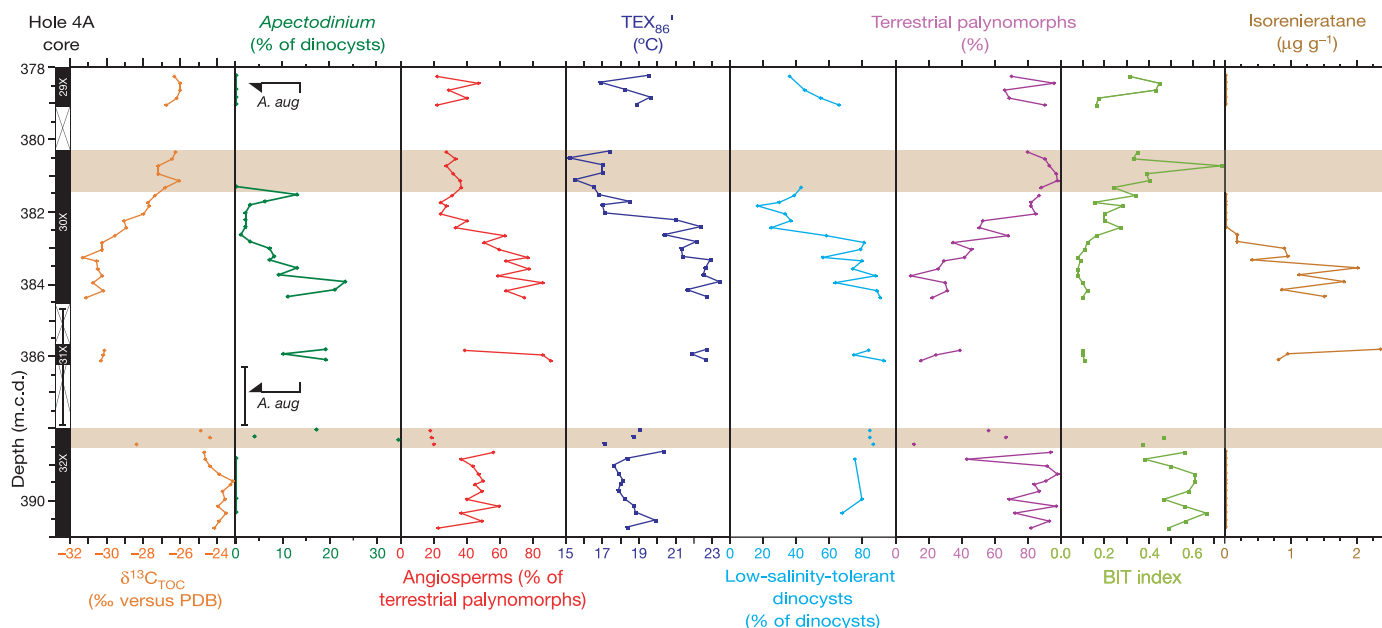


Figure 2 | Core recovery and palynological and geochemical results across the PETM of IODP Hole 302-4A. Core 31X was plotted 100 cm lower than m.c.d.⁵ for illustration purposes. Error bars connected to Core 31X in the recovery column indicate the uncertainty of its stratigraphic position (see Supplementary Information). Orange bars indicate intervals affected by drilling disturbance. Stable carbon isotopes are expressed relative to the

PeeDee Belemnite standard. Low-salinity-tolerant dinocysts comprise *Senegalinium* spp., *Cerodinium* spp., and *Polysphaeridium* spp., while *Membranosphaera* spp., *Spiniferites ramosus* complex, and *Areoligera-Glaphyrocysta* cpx. represent the typical normal marine species²⁵ (Supplementary Fig. S-1). Arrows and *A. aug* indicate the first and last occurrences of dinocyst *Apectodinium augustum*.

PETM to a sea level rise, an interpretation consistent with evidence from other neritic locations²⁷. The gradual return to more terrestrial influence later in the event probably results from subsequent regression. Despite the transgression, low-salinity-tolerant dinocysts remain dominant (Fig. 2), showing that brackish surface waters persisted during the PETM. If the earliest Palaeogene greenhouse world was continental-ice-free, a mechanism for the sea-level rise may comprise the ~5 m thermal expansion of sea water expected from a 5–8 °C (refs 2, 3) increase in deep ocean temperatures.

The occurrence of laminated sediments from the onset of the PETM (although hardly visible in Core 31X due to its disturbed state) up to 382.5 m.c.d., and the absence of benthic foraminiferal linings (Supplementary Fig. S-3A) suggest that bottom waters became anoxic during the PETM. Within the laminated interval, derivatives of the characteristic pigment isorenieratene are recorded in concentrations of up to $2 \mu\text{g g}^{-1}$ sediment; they are below the detection limit outside this interval (Fig. 2; Supplementary Fig. S-3). These compounds are derived from the brown strain of photosynthetic green sulphur bacteria, which requires euxinic (anoxic and sulphidic) conditions to thrive²⁸. Accordingly, at the PETM photic zone euxinia developed at the drill site coincident with bottom water anoxia, which gradually disappeared during the recovery of SST and $\delta^{13}\text{C}$ excursion (Fig. 2).

We can exclude selective preservation as a mechanism to explain the marked changes in organic biomarkers and palynomorph assemblages that occur in coincidence with water column anoxia. First, the preservation of organic matter is also excellent outside the laminated interval (>2% TOC on average; Supplementary Fig. S-3A) and second, most of our proxies compare the relative abundance of structurally similar organic compounds that are equally susceptible to oxidation (see Supplementary Information).

The euxinic conditions were potentially caused by several factors. For example, increased freshwater input, greater nutrient load and warmer temperatures would all conspire to reduce dissolved O_2 in the water column. However, given the shallow water depth of the site, an important factor was probably intense stratification due to the influence of a brackish surface-water lid. Although several mechanisms could drive such stratification, given that low-salinity-tolerant dinocysts remain dominant despite the more distal position of the site, and given the data presented in a companion paper¹⁶, the simplest explanation is that decreased mixing resulted from increased SSTs and enhanced fluvial runoff, with the latter also supplying extra nutrients to increase production and saturate photic zone respiration. The termination of euxinic conditions coincides with increasing surface salinities¹⁶ (Fig. 2) and cooling, suggesting an increase of mixing with more-saline deeper waters.

Even if we assume that our TEX_{86} temperatures represent summer values (see Supplementary Information), palaeoclimate models simulating the early Palaeogene world with 2,000 p.p.m.v. of CO_2 in the atmosphere⁶ underestimate Arctic Ocean summer SSTs by at least 15 °C for the PETM and 10 °C for the surrounding late Palaeocene and early Eocene. This discrepancy may be even larger because the initial part of the PETM, and potentially also the strata formed under maximum temperatures, were possibly not recovered (Fig. 2). On the other hand, the magnitude of the carbon isotope excursion is comparable to previous studies and peak PETM temperatures lagged the onset of the carbon isotope excursion by ~40 kyr (ref. 1), indicating that optimum Arctic SSTs are probably covered in our record.

The models consistently predict pole-to-Equator temperature gradients of ~30 °C (ref. 29). Such gradients represent significant overestimates because they would imply unrealistically warm tropical SSTs considering our polar temperatures. The high polar temperatures and reduced pole-to-Equator temperature gradients cannot be explained by invoking even greater greenhouse gas concentrations because this would elevate tropical SSTs, which in existing model predictions already match or exceed those determined from proxy

records at low-latitude locations⁶. Also, ocean heat transport is unlikely to be the cause because this requires a threefold increase, which cannot be simulated in the current generation of fully coupled ocean-atmosphere climate models²⁹. Similarly, atmospheric general circulation models do not support strong enough positive feedbacks in atmospheric heat transport³⁰. Consequently, we surmise that physical processes that are not yet incorporated in the models operated in conjunction with high greenhouse gas concentrations to enhance polar warmth and reduce the pole-to-Equator temperature gradient during the early Palaeogene. These processes could include high-latitude warming and tropical cooling through the enhancement of polar stratospheric clouds⁷, and hurricane-induced ocean mixing⁸, respectively.

With latest Palaeocene SSTs of 18 °C it is not likely that ice was present in the Arctic. This implies that the PETM at Hole 4A allows us to examine the Arctic environment and the nature of polar amplification during a time of massive greenhouse gas emissions and extreme global warming in the absence of ice-albedo feedbacks. Interestingly, polar amplification of temperature rise at the PETM appears to have been minor (Fig. 2; refs 1–3), suggesting that the strengthening of the mechanism that caused above early Palaeogene polar temperature amplification was small at the PETM. Our extremely warm polar temperatures indicate that, despite much recent progress, feedbacks responsible for early Palaeogene mid- to high-latitude warmth remain poorly understood and are not implemented in existing climate models.

METHODS

Palynology. Sediments were oven-dried at 60 °C. To ~2 g of sediment, a known amount of *Lycopodium* spores were added, after which the sample was treated with 30% HCl and twice with 30% HF for carbonate and silicate removal, respectively. After sieving over a 15- μm nylon mesh sieve, residues were analysed at 500 $\times$ magnification until a minimum of 200 dinocysts had been recorded per sample. Absolute quantitative numbers were calculated using the relative number of *Lycopodium*.

Organic geochemistry. Powdered and freeze-dried sediments were analysed for %TOC and $\delta^{13}\text{C}_{\text{TOC}}$ with a Fison NA 1500 CNS analyser, connected to a Finnigan Delta Plus mass spectrometer. Analytical precision and accuracy were determined by replicate analyses and by comparison with international and in-house standards, and were better than 0.1% and 0.1‰ for %TOC and $\delta^{13}\text{C}_{\text{TOC}}$, respectively.

Powdered and freeze-dried sediments (1–3 g dry mass) were extracted with dichloromethane (DCM)/methanol (2:1) by using the Dionex accelerated solvent extraction technique. The extracts were separated by Al_2O_3 column chromatography using hexane/DCM (9:1) and DCM/methanol (1:1) to yield the apolar and polar fractions, respectively. The apolar fractions were analysed for isorenieratene derivatives by gas chromatography and gas chromatography/mass spectrometry, while the polar fractions were analysed for tetraether lipids and used to calculate TEX_{86} (see Supplementary Information; reproducibility was within ~1 °C) and BIT (see ref. 26 for method description) indices.

Received 7 July 2005; accepted 21 February 2006.

1. Zachos, J. C. *et al.* A transient rise in tropical sea surface temperature during the Paleocene-Eocene thermal maximum. *Science* **302**, 1151–1154 (2003).
2. Kennett, J. P. & Stott, L. D. Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Palaeocene. *Nature* **353**, 225–229 (1991).
3. Tripati, A. & Elderfield, H. Deep-sea temperature and circulation changes at the Paleocene-Eocene thermal maximum. *Science* **308**, 1894–1898 (2005).
4. Dickens, G. R., O'Neil, J. R., Rea, D. K. & Owen, R. M. Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Paleocene. *Paleoceanography* **10**, 965–971 (1995).
5. Backman, J., Moran, K., McInroy, D. B., Mayer, L. A. & the Expedition 302 Scientists. Arctic Coring Expedition (ACEX). *Proc. ODP* **302** 1/doi:10.2204/iodp.proc.302.2006 (Integrated Ocean Drilling Program Management International, College Station, Texas, 2006).
6. Shellito, C. J., Sloan, L. C. & Huber, M. Climate model sensitivity to atmospheric CO_2 levels in the Early-Middle Paleogene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **193**, 113–123 (2003).
7. Sloan, L. C. & Pollard, D. Polar stratospheric clouds: A high latitude warming mechanism in an ancient greenhouse world. *Geophys. Res. Lett.* **25**, 3517–3520 (1998).

8. Emanuel, K., DesAutels, C., Holloway, C. & Korty, R. Environmental Control of Tropical Cyclone Intensity. *J. Atmos. Sci.* **61**, 843–858 (2004).
9. Koch, P. L., Zachos, J. C. & Gingerich, P. D. Correlation between isotope records in marine and continental carbon reservoirs near the Palaeocene/Eocene boundary. *Nature* **358**, 319–322 (1992).
10. Röhl, U., Bralower, T. J., Norris, G. & Wefer, G. A new chronology for the late Paleocene thermal maximum and its environmental implications. *Geology* **28**, 927–930 (2000).
11. Crouch, E. M. *et al.* Global dinoflagellate event associated with the late Paleocene thermal maximum. *Geology* **29**, 315–318 (2001).
12. Thomas, E. & Shackleton, N. J. in *Correlation of the Early Paleogene in Northwestern Europe* (eds Knox, R. W. O. B., Corfield, R. M. & Dunay, R. E.) 401–441 (Geological Society of London Special Publication 101, 1996).
13. Wing, S. L. in *Late Paleocene–early Eocene Climatic and Biotic Events in the Marine and Terrestrial Records* (eds Aubry, M.-P., Lucas, S. G. & Berggren, W. A.) 380–400 (Columbia Univ. Press, New York, 1998).
14. Bowen, G. J. *et al.* Mammalian dispersal at the Paleocene/Eocene boundary. *Science* **295**, 2062–2065 (2002).
15. Bujak, J. P. & Brinkhuis, H. in *Late Paleocene–early Eocene Climatic and Biotic Events in the Marine and Terrestrial Records* (eds Aubry, M.-P., Lucas, S. G. & Berggren, W. A.) 277–295 (Columbia Univ. Press, New York, 1998).
16. Pagani, M. *et al.* Arctic's hydrology during global warming at the Paleocene/Eocene thermal maximum. *Nature* (submitted).
17. Steurbaut, E. *et al.* in *Causes and Consequences of Globally Warm Climates in the Early Paleogene* (eds Wing, S. L., Gingerich, P., Schmitz, B. & Thomas, E.) 291–317 (Geological Society of America Special Paper 369, Boulder, Colorado, 2003).
18. Schouten, S., Hopmans, E. C., Schefuß, E. & Sinninghe Damsté, J. S. Distributional variations in marine crenarchaeotal membrane lipids: a new tool for reconstructing ancient sea water temperatures? *Earth Planet. Sci. Lett.* **204**, 265–274 (2002).
19. Wuchter, C., Schouten, S., Coolen, M. J. L. & Sinninghe Damsté, J. S. Temperature-dependent variation in the distribution of tetraether membrane lipids of marine Crenarchaeota: Implications for TEX₈₆ paleothermometry. *Paleoceanography* **19**, doi:10.1029/2004PA001041 (2004).
20. Markwick, P. J. Fossil crocodilians as indicators of Late Cretaceous and Cenozoic climates: Implications for using palaeontological data in reconstructing palaeoclimate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **137**, 205–271 (1998).
21. Tripathi, A., Zachos, J., Marincovich, L. Jr & Bice, K. Late Paleocene Arctic coastal climate inferred from molluscan stable and radiogenic isotope ratios. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **170**, 101–113 (2001).
22. Jenkyns, H. C., Forster, A., Schouten, S. & Sinninghe Damsté, J. S. High temperatures in the Late Cretaceous Arctic Ocean. *Nature* **432**, 888–892 (2004).
23. Fricke, H. C. & Wing, S. L. Oxygen isotope and paleobotanical estimates of temperature and $\delta^{18}\text{O}$ -latitude gradients over North America during the early Eocene. *Am. J. Sci.* **304**, 612–635 (2004).
24. Brinkhuis, H. *et al.* Episodic fresh surface waters in the Eocene Arctic Ocean. *Nature* doi:10.1038/nature04692 (this issue).
25. Sluijs, A., Pross, J. & Brinkhuis, H. From greenhouse to icehouse; organic-walled dinoflagellate cysts as paleoenvironmental indicators in the Paleogene. *Earth Sci. Rev.* **68**, 281–315 (2005).
26. Hopmans, E. C. *et al.* A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids. *Earth Planet. Sci. Lett.* **224**, 107–116 (2004).
27. Speijer, R. P. & Morsi, A.-M. M. Ostracode turnover and sea-level changes associated with the Paleocene-Eocene thermal maximum. *Geology* **30**, 23–26 (2002).
28. Sinninghe-Damsté, J. S., Wakeham, S. G., Kohnen, M. E. L., Hayes, J. M. & De Leeuw, J. W. A 6,000-year sedimentary molecular record of chemocline excursions in the Black Sea. *Nature* **362**, 827–829 (1993).
29. Huber, M., Sloan, L. C. & Shellito, C. J. in *Causes and Consequences of Globally Warm Climates in the Early Paleogene* (eds Wing, S. L., Gingerich, P. D., Schmitz, B. & Thomas, E.) 25–47 (Geological Society of America Special Paper 369, Boulder, Colorado, 2003).
30. Caballero, R. & Langen, P. L. The dynamic range of poleward energy transport in an atmospheric general circulation model. *Geophys. Res. Lett.* **32**, doi: 10.1029/2004GL021581 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.S. thanks the Utrecht Biogeology Centre for funding. H.B. thanks the Netherlands Organization for Scientific Research, and Utrecht University for enabling participation in the ACEX expedition. M.H. thanks the Purdue Climate Change Research Center, ITaP and the Purdue Research Foundation for their continued support. This research used samples and data provided by the IODP. We thank L. Bik, J. van Tongeren, N. Welters and A. van Dijk for technical support, and C. E. Stickley for discussions.

Author Contributions A.S. and H.B. carried out the palynology, A.S. & G.-J.R. the $\delta^{13}\text{C}_{\text{TOC}}$, S.S., M.W. and J.S.S.D. the TEX₈₆, BIT and isorenieratane analyses. R.S. generated the hydrogen index data. J.B. and K.M. were the co-chiefs of the ACEX expedition. N.P., J.M. and the Expedition 302 Scientists were involved in generating shipboard and shore-based ACEX data. A.S., S.S., M.P., H.B., J.S.S.D., G.R.D., M.H. and L.J.L. contributed to interpreting the data and writing the paper.

Author Information Reprints and permissions information is available at ngp.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.S. (a.sluijs@bio.uu.nl) or S.S. (schouten@nioz.nl)

Expedition 302 Scientists (those not listed above): Steve Clemens¹, Thomas Cronin², Frédérique Eynaud³, Jérôme Gattacceca⁴, Martin Jakobsson⁵, Ric Jordan⁶, Michael Kaminski⁷, John King⁸, Nalân Koc⁹, Nahysa C. Martinez¹⁰, David McInroy¹¹, Theodore C. Moore Jr¹², Matthew O'Regan⁸, Jonaotaro Onodera¹³, Heiko Pälike¹⁴, Brice Rea¹⁵, Domenico Rio¹⁶, Tatsuhiko Sakamoto¹⁷, David C. Smith⁸, Kristen E. K. St John¹⁸, Itsuki Suto¹⁹, Noritoshi Suzuki²⁰, Kozo Takahashi¹³, Mahito Watanabe²¹ & Masanobu Yamamoto²²

¹Geological Sciences, Brown University, 324 Brook Street, PO Box 1846, Providence, Rhode Island 02912-1846, USA. ²US Geological Survey, Eastern Earth Surface Processes Team, 926A USGS National Center, Reston, Virginia 20192, USA. ³Department de Géologie et Océanographie, Université Bordeaux 1, Avenue des facultés, c/o Bernei Housen, 33405 Talence Cedex, France. ⁴Department of Geophysics, CEREGE (CNRS)/University of Aix-Marseille 3, BP 80, 13545 Aix-en-Provence Cedex 4, France. ⁵Department of Geology and Geochemistry, Stockholm University, 106 91 Stockholm, Sweden. ⁶Department of Earth and Environmental Sciences, Faculty of Science, Yamagata University, 1-4-12 Kojirakawa-machi, Yamagata 990-8560, Japan. ⁷Department of Earth Sciences, University College London, Gower Street, London WC1E 6BT, UK. ⁸Graduate School of Oceanography, University of Rhode Island, Narragansett Bay Campus, South Ferry Road, Narragansett, Rhode Island 02882, USA. ⁹Norwegian Polar Institute, Polar Environmental Center, N-9296 Tromsø, Norway. ¹⁰Department of Earth Sciences, Boston University, 685 Commonwealth Avenue, Boston, Massachusetts 02215, USA. ¹¹British Geological Survey, Murchison House, West Mains Road, Edinburgh, EH9 3LA, UK. ¹²Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109-1063, USA. ¹³Department of Earth and Planetary Sciences, Graduate School of Sciences, Kyushu University, Hakozaki 6-10-1, Higashi-ku, Fukuoka 812-8581, Japan. ¹⁴School of Ocean and Earth Science, University of Southampton, Southampton Oceanography Center, European Way, Southampton SO14 3ZH, UK. ¹⁵Department of Geography and Environment, School of Geosciences, University of Aberdeen, Elphinstone Road, Aberdeen AB24 3UF, UK. ¹⁶Department of Geology, Paleontology and Geophysics, University of Padova, Via Giotto 1 I-35137 Padova, Italy. ¹⁷Institute for Research on Earth Evolution (IFREE), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Natsushima-cho 2-15, Yokosuka 237-0061, Japan. ¹⁸Department of Geology and Environmental Science, MSC 7703, James Madison University, Harrisonburg, Virginia 22807, USA. ¹⁹Institute of Life and Environmental Science, University of Tsukuba, Tennoudai 1-1-1, Tsukuba, Ibaraki 305-8572, Japan. ²⁰Institute of Geology and Paleontology, Graduate School of Science, Tohoku University, Aramaki, Aoba, Aoba-ku, Sendai City 980-8578, Japan. ²¹Institute of Geoscience, National Institute of Advanced Industrial Science, and Technology (Geological Survey of Japan) AIST Tsukuba Central 7, Higashi-1-1-1, Tsukuba, Ibaraki 305-8567, Japan. ²²Graduate School of Environmental Earth Science, Hokkaido University, Kita-10, Nishi-5, Kita-ku, Sapporo 060-0810, Japan

LETTERS

Diapir-induced reorientation of Saturn's moon Enceladus

Francis Nimmo¹ & Robert T. Pappalardo^{2†}

Enceladus is a small icy satellite of Saturn. Its south polar region consists of young, tectonically deformed terrain and has an anomalously high heat flux^{1,2}. This heat flux is probably due to localized tidal dissipation within either the ice shell³ or the underlying silicate core⁴. The surface deformation is plausibly due to upwelling of low-density material (diapirism⁵) as a result of this tidal heating. Here we show that the current polar location of the hotspot can be explained by reorientation of the satellite's rotation axis because of the presence of a low-density diapir. If the diapir is in the ice shell, then the shell must be relatively thick and maintain significant rigidity (elastic thickness greater than ~ 0.5 km); if the diapir is in the silicate core, then Enceladus cannot possess a global subsurface ocean, because the core must be coupled to the overlying ice for reorientation to occur. The reorientation generates large (~ 10 MPa) tectonic stress patterns⁶ that are compatible with the observed deformation of the south polar region². We predict that the distribution of impact craters on the surface will not show the usual leading hemisphere–trailing hemisphere asymmetry. A low-density diapir also yields a potentially observable negative gravity anomaly.

Enceladus (radius 252 km) is near the 2:1 mean motion resonance with Dione, forcing its orbital eccentricity to 0.0047. Tidal heating^{3,7}, possibly including either resonant libration⁸ or interactions with the satellite Janus⁹, is probably sufficient to cause geological activity; the maximum likely heat output is of the order of 10^3 GW (ref. 3). This tidal deformation may also lead to localized shear heating^{2,10} along active polar tectonic features observed by the Cassini spacecraft.

The ~ 300 -km-wide, tectonically deformed southern region of Enceladus², and other older deformed regions, show some resemblance to the three tectonically deformed 'coronae' on the similarly small (236 km radius) uranian satellite Miranda^{5,11}. Miranda's coronae are analogously ovoidal and large (~ 200 – 300 km across), are tectonically deformed with distinct inner and outer deformation zones, and are located along the satellite's larger principal axes of inertia. They are generally inferred to be regions of tectonic deformation and icy volcanism above large-scale ice diapirs^{5,11}. The location of the coronae suggests that true polar wander occurred¹², and past reorientation is consistent with spatial variations in the distribution of fresh craters on Miranda¹³. On Earth and Mars, convective upwellings are likely to have resulted in true polar wander in the past^{14,15}, while ice shell thickness variations may have caused similar effects on Jupiter's moon Europa¹⁶.

Tidal heating within the silicate core of Enceladus⁴ may produce warm silicate upwellings there, and could also lead to diapirism in a sufficiently thick overlying ice shell. Alternatively, localized heating within the ice shell itself may cause a low-density diapir¹⁷. Below, we consider the circumstances under which a large, low-density diapir in either the silicate or the ice portion of Enceladus could have caused

reorientation of the satellite, moving the diapiric region towards the satellite's spin axis.

We first consider an ice diapir, as depicted in Fig. 1. If the overlying lithosphere has negligible rigidity, the mass deficit at depth will be compensated by a surface mass excess, generated by upwarped topography. Because that mass excess is closer to the surface than the interior mass deficit, the net effect is to generate a positive gravitational potential anomaly, which would tend to reorient the region towards the equator. In contrast, for the case of an infinitely rigid lithosphere, there would be no surface topography, thus producing a net negative potential anomaly, tending to reorient the region towards the pole. A subsurface diapir therefore can result in either poleward or equatorward reorientation, depending on the rigidity (or elastic thickness) of the lithosphere¹⁶. Reorientation is opposed by the frozen-in component of the triaxial satellite's tidal and rotational bulges^{18,19}; thus, other things being equal, a satellite with lower rigidity will have smaller permanent bulges and is more likely to undergo reorientation.

Reorientation is favoured if density anomalies persist for times long compared to the viscous relaxation timescale^{18,20}. This condition

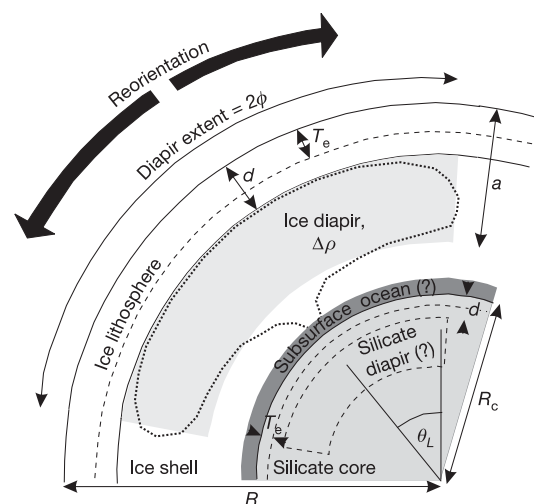


Figure 1 | Schematic diagram of subsurface diapirism on Enceladus. The light shaded area approximates the shape of an inferred diapir (dotted line) in the ice shell, and has a density contrast $\Delta\rho$ with the surrounding ice. The dashed region within the darker shaded area denotes an alternative diapir location within the silicate core. For ice and silicate densities of 950 kg m^{-3} and $3,500 \text{ kg m}^{-3}$, respectively, $R_c = 161$ km using a bulk density for Enceladus² of $1,610 \text{ kg m}^{-3}$. For the ice diapir we use $a = 91$ km; for the silicate diapir $a = 155$ km, and d and T_e are measured from the surface of the silicate core. In both cases, $\phi = 35^\circ$ and the initial load colatitude $\theta_L = 45^\circ$.

¹Department of Earth Sciences, University of California Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA. ²Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, Colorado 80309-0392, USA. [†]Present address: Earth and Space Sciences Division, Jet Propulsion Laboratory, Pasadena, California 91109, USA.

is likely to be satisfied on Enceladus for both the ice shell¹⁶ and silicate core¹⁴. Following the approach of Matsuyama *et al.*¹⁸, for a synchronously rotating satellite, the long-term angular reorientation δ due to an imposed potential anomaly is given by

$$\delta = \frac{1}{2} \tan^{-1} \left(\frac{Q^* \sin 2\theta_L}{n - Q^* \cos 2\theta_L} \right) \quad (1)$$

where θ_L is the initial colatitude of the centre of the anomaly, n is a parameter that depends on the longitudinal location of the anomaly relative to the tidal bulge (see Supplementary Information), and Q^* is a parameter that describes both the size of the potential anomaly and the non-hydrostatic component of rotational flattening, which opposes reorientation.

The quantity Q^* is given¹⁸ by $Q^* = 3\sqrt{5}G_{20}/R^2\Omega^2(k_2^f - k_2)$ where G_{20} is the second-degree potential anomaly due to the diapir, R is the mean satellite radius, Ω is the rotation angular frequency of the satellite, and k_2^f and k_2 are the degree-two tidal Love numbers for the satellite without and with an elastic layer, respectively¹⁸. Smaller values of k_2 result in a larger permanent bulge and less reorientation, while larger potential anomalies result in greater reorientation. Here we calculate k_2 using a multi-layered Maxwell viscoelastic model²¹ in which all regions except the elastic part of the lithosphere are assumed to be inviscid (Supplementary Information).

The second-degree potential anomaly G_{20} due to a partially compensated load depends on the diapir angular half-width ϕ and radial extent $a - d$ (see Fig. 1), the density contrast between the diapir and the surrounding material $\Delta\rho$, and the degree of compensation C , and is derived in the Supplementary Information. C depends on the ability of the cold elastic part of the ice or silicate layer (thickness T_e ; see Fig. 1) to resist deformation, and is calculated using the method of ref. 22.

Figure 2a shows how Q^* , C and k_2 vary as a function of lithospheric thickness d , where, following ref. 23, the elastic thickness $T_e = 0.4d$ above either silicate or ice diapirs. For the ice diapir (solid lines), the degree of compensation C decreases as d increases. At some critical degree of compensation (~ 0.8 in this case) the potential anomaly switches from positive to negative, and as a result so does Q^* . This indicates that at greater elastic thicknesses, the diapir will reorient towards the pole. The solid lines in Fig. 2b show the angular reorientation (equation (1)) produced by an ice diapir. Lower load latitudes generally lead to greater reorientation, and larger density contrasts lead to greater reorientation. Low values of d lead to equatorward motion, while d values in excess of ~ 1.3 km lead to poleward reorientation.

The solid curves in Fig. 2 demonstrate that an ice diapir within a thick shell and with a sufficiently large density contrast can lead to significant poleward reorientation (up to 30°), if the elastic thickness exceeds ~ 0.5 km. Thus, if an ice diapir has reoriented Enceladus, the polar location of the active region suggests that the satellite has retained some near-surface rigidity, despite the observed surface deformation. An elastic thickness of ~ 0.5 km is similar to results obtained for Ganymede²³, which apparently underwent an ancient episode of tidal deformation and heating. Similarly, the elastic thickness of Miranda in regions adjacent to the coronae was ~ 2 km at the time of deformation⁵.

Figure 2 also shows that large density contrasts are required to achieve significant reorientation, suggesting that an ice diapir must be primarily compositional, rather than thermal, in origin. Partial melting of a tidally heated diapir will preferentially remove low-melting-temperature, dense components, such as salts, leading to compositional buoyancy^{17,24}.

The reorientation due to a diapir within the silicate core (dashed lines in Fig. 2) is comparable to that for an ice diapir, and promotes poleward reorientation. A co-axial combination of silicate and ice diapirism would lead to larger angular reorientations, because the reorientation depends on the total potential anomaly.

Observations of the tectonic deformation around the south pole

of Enceladus² support the reorientation hypothesis. The observed orientations of polar contractional and lower-latitude extensional structures are consistent with the stress patterns generated by moderate true polar wander⁶. Strike-slip activity in subpolar latitudes is also predicted, and will provide an additional observational test of the hypothesis. If the flattening of Enceladus is close to hydrostatic², a 30° reorientation will give rise to stresses approaching 10 MPa (ref. 6), easily enough to generate significant deformation.

We can think of at least two additional tests. First, the distribution of impact craters—expected to show leading–trailing hemisphere asymmetry if the satellite orbits synchronously—will be affected by, and could constrain, reorientation¹³. Second, Fig. 2 shows that a negative, long-wavelength gravity anomaly (with an amplitude of a few mGal at 200 km spacecraft altitude) will be associated with the anomalous south polar region. Regional gravity anomalies similar to this have been detected at Ganymede by the Galileo spacecraft²⁵.

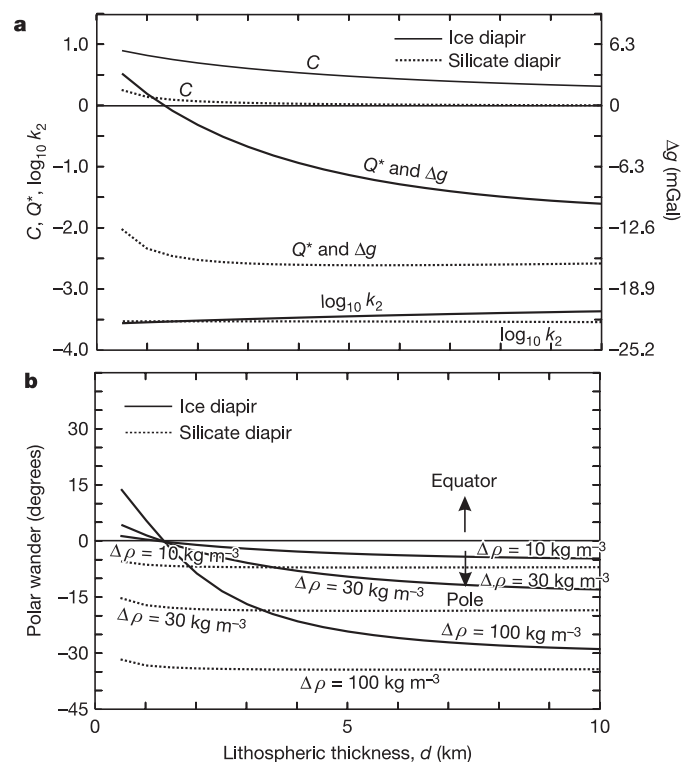


Figure 2 | Variation of parameters affecting reorientation, and the degree of reorientation itself, as a function of d , the lithospheric thickness. **a**, C is the compensation factor, Q^* is the reorientation factor defined in the text, and k_2 is the Love number. Love numbers are calculated using a multi-layer Maxwell viscoelastic code²¹ (see Supplementary Information). Other parameters are defined in Fig. 1 legend. The elastic thickness $T_e = 0.4d$, the fluid Love number $k_2^f = 0.75$ and the satellite rotation frequency $\Omega = 5.31 \times 10^{-5} \text{ s}^{-1}$. Solid lines denote ice diapir; dashed lines denote silicate diapir. For the ice diapir case the silicate core is assumed to behave elastically, resulting in a reduced k_2 . Importantly, because $k_2 \ll k_2^f$ the denominator of Q^* is almost independent of k_2 and thus neither Q^* nor the amount of reorientation are very sensitive to the value of k_2 adopted. The quantity Δg is the degree-2 gravity anomaly in mGal (right-hand scale), which is proportional to Q^* and is derived by assuming a spacecraft altitude of 200 km (see Supplementary Information). Q^* and Δg are calculated using $\Delta\rho = 100 \text{ kg m}^{-3}$. **b**, Angular reorientation δ (equation (1)) as a function of d for different density contrasts. Positive values of δ indicate equatorward reorientation; negative values indicate poleward reorientation. We use $\theta_L = 45^\circ$, $\phi = 35^\circ$, $n = 1$. The reorientation for a silicate diapir is much less sensitive to the elastic thickness assumed than that for an ice diapir, because the higher rigidity of silicates compared to ices means that the compensation factor C is close to zero for the range of T_e values that we considered (Fig. 2a).

We can also test whether the inferred reorientation is due to diapirism within the ice shell, the silicate core, or indeed some combination of the two. First, reorientation due to an ice diapir requires an elastic thickness in excess of ~ 0.5 km. This requirement can be compared with T_e values inferred using local topographic measurements²³. Second, if diapirism in the silicate core is primarily responsible for the inferred reorientation, the ice shell must be efficiently coupled to the rocky core beneath. The effect of a global subsurface ocean would be to decouple the shell from the underlying material¹⁶; thus, if a silicate diapir is responsible for reorientation, Enceladus must lack a global subsurface ocean. Subsurface oceans have been detected in several galilean satellites, but detecting an internal ocean using magnetometry in the Saturn system is significantly more challenging.

There are several additional effects that we have not included. We have neglected both the inhibiting effect of dissipation caused by the reorientation itself¹⁶, and possible spatial variations in T_e . Enceladus' orbital eccentricity, and thus the amount of tidal heating, is unlikely to have been constant through time²⁶. This raises the possibility that more than one diapir may have arisen, leading to successive reorientations of the body and a complicated (but potentially decipherable) deformation history. Similarly, the dimensions and density contrast of each rising diapir, and the elastic thickness of the overlying rigid layer, are likely to have changed with time, generating more complicated and potentially non-monotonic reorientation paths^{14,15,18,20}. However, our main conclusions are unlikely to be affected by these details.

The principal requirements for reorientation to occur are the development of sufficiently large and long-lived density contrasts of sufficient lateral and radial extent. Unless there are large vertical viscosity gradients²⁷, vigorous convection is unlikely to generate such contrasts, because both the characteristic length scales and timescales of convective features tend to be small at high Rayleigh numbers¹⁴. Conversely, the initiation of either convection or a Rayleigh–Taylor upwelling will occur on a length scale that is usually comparable to the fluid layer depth²⁸. Thus, bodies which are, or once were, active enough to generate density contrasts, but not to convect vigorously, are most likely to have undergone reorientation. The mid-sized icy satellites, including those of Saturn, have in many cases undergone geological deformation in the past²⁹, and are expected to be close to the onset of convection³⁰. We therefore anticipate possible additional examples of mid-sized satellite reorientation as more Cassini data are returned.

Received 15 March; accepted 10 April 2006.

- Spencer, J. R. *et al.* Cassini encounters Enceladus: Background and the discovery of a south polar hot spot. *Science* **311**, 1401–1405 (2006).
- Porco, C. C. *et al.* Cassini observes the active south pole of Enceladus. *Science* **311**, 1393–1401 (2006).
- Ross, M. N. & Schubert, G. Viscoelastic models of tidal heating in Enceladus. *Icarus* **78**, 90–101 (1989).
- Matson, D. L. *et al.* Enceladus' interior and geysers—possibility for hydrothermal geochemistry. *Lunar Planet. Sci. Conf. XXXVII*, abstr. 2219 (2006).
- Pappalardo, R. T., Reynolds, S. J. & Greeley, R. Extensional tilt blocks on Miranda: Evidence for an upwelling origin of Arden Corona. *J. Geophys. Res.* **102**, 13369–13379 (1997).
- Melosh, H. J. Tectonic patterns on a reoriented planet: Mars. *Icarus* **44**, 745–751 (1980).
- Squyres, S. W., Reynolds, R. T. & Cassen, P. M. The evolution of Enceladus. *Icarus* **53**, 319–331 (1983).
- Wisdom, J. Spin-orbit secondary resonance dynamics of Enceladus. *Astron. J.* **128**, 484–491 (2004).
- Lissauer, J. J., Peale, S. J. & Cuzzi, J. N. Ring torque on Janus and the melting of Enceladus. *Icarus* **58**, 159–168 (1984).
- Gaidos, E. & Nimmo, F. Tectonics and water on Europa. *Nature* **405**, 637 (2000).
- Greenberg, R., *et al.* in *Uranus* (eds Bergstrahl, J. T. *et al.*) 693–735 (Univ. Arizona Press, Tucson, 1991).
- Janes, D. M. & Melosh, H. J. Sinkers tectonics—an approach to the surface of Miranda. *J. Geophys. Res.* **93**, 3127–3143 (1988).
- Plescia, J. B. Cratering history of Miranda—implications for geologic processes. *Icarus* **73**, 442–461 (1988).
- Richards, M. A., Bunge, H. P., Ricard, Y. & Baumgardner, J. R. Polar wandering in mantle convection models. *Geophys. Res. Lett.* **26**, 1777–1780 (1999).
- Moser, J., Yuen, D. A., Larsen, T. B. & Matyska, C. Dynamical influences of depth-dependent properties on mantle upwellings and temporal variations of the moment of inertia. *Phys. Earth Planet. Inter.* **102**, 153–170 (1997).
- Ojakangas, G. W. & Stevenson, D. J. Polar wander of an ice shell on Europa. *Icarus* **81**, 242–270 (1989).
- Sotin, C., Head, J. W. & Tobie, G. Europa: tidal heating of upwelling thermal plumes and the origin of lenticulae and chaos melting. *Geophys. Res. Lett.* **29**, 1233, doi:10.1029/2001GL013844 (2002).
- Matsuyama, I., Mitrovica, J. X., Manga, M., Perron, J. T. & Richards, M. A. Rotational stability of dynamic planets with elastic lithospheres. *J. Geophys. Res.* **111**, E02003, doi:10.1029/2005JE002447 (2006).
- Willemann, R. J. Reorientation of planets with elastic lithospheres. *Icarus* **60**, 701–709 (1984).
- Tsai, V. C. & Stevenson, D. J. Theoretical constraints on true polar wander. *Eos* **86**, GP21A–02 (2005).
- Moore, W. B. & Schubert, G. The tidal response of Europa. *Icarus* **147**, 317–319 (2000).
- Turcotte, D. L., Willemann, R. J., Haxby, W. F. & Norberry, J. Role of membrane stresses in the support of planetary topography. *J. Geophys. Res.* **86**, 3951–3959 (1981).
- Nimmo, F., Pappalardo, R. T. & Giese, B. Elastic thickness and heat flux estimates on Ganymede. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL013976 (2002).
- Nimmo, F., Pappalardo, R. T. & Giese, B. On the origins of band topography Europa. *Icarus* **166**, 21–32 (2003).
- Palguta, J., Anderson, J. D., Schubert, G. & Moore, W. B. Mass anomalies on Ganymede. *Icarus* **180**, 428–441 (2006).
- Ojakangas, G. W. & Stevenson, D. J. Episodic volcanism of tidally-heated satellites with application to Io. *Icarus* **66**, 341–358 (1986).
- McNamara, A. K. & Zhong, S. J. Degree-one mantle convection: Dependence on internal heating and temperature-dependent rheology. *Geophys. Res. Lett.* **32**, L01301, doi:10.1029/2004GL021082 (2005).
- Turcotte, D. L. & Schubert, G. *Geodynamics* 2nd edn (Cambridge Univ. Press, Cambridge, UK, 2002).
- Morrison, D., Owen, T. & Soderblom, L. A. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 764–801 (Univ. Arizona Press, Tucson, 1986).
- McKinnon, W. B. in *Solar System Ices* (eds Schmitt, B., de Bergh, C. & Festou, M.) 525–550 (Kluwer Academic, Dordrecht, 1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Moore, J. Melosh and M. Mullen for comments. This research was supported by NASA PGG and OPR.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to F.N. (fnimmo@es.ucsc.edu).

Dimensional reduction at a quantum critical point

S. E. Sebastian¹, N. Harrison², C. D. Batista³, L. Balicas⁴, M. Jaime², P. A. Sharma², N. Kawashima⁵ & I. R. Fisher¹

Competition between electronic ground states near a quantum critical point^{1,2} (QCP)—the location of a zero-temperature phase transition driven solely by quantum-mechanical fluctuations—is expected to lead to unconventional behaviour in low-dimensional systems³. New electronic phases of matter have been predicted to occur in the vicinity of a QCP by two-dimensional theories^{3–8}, and explanations based on these ideas have been proposed for significant unsolved problems in condensed-matter physics, such as non-Fermi-liquid behaviour and high-temperature superconductivity. But the real materials to which these ideas have been applied are usually rendered three-dimensional by a finite electronic coupling between their component layers; a two-dimensional QCP has not been experimentally observed in any bulk three-dimensional system, and mechanisms for dimensional reduction have remained the subject of theoretical conjecture^{9–11}. Here we show evidence that the Bose–Einstein condensate of spin triplets in the three-dimensional Mott insulator $\text{BaCuSi}_2\text{O}_6$ (refs 12–16) provides an experimentally verifiable example of dimensional reduction at a QCP. The interplay of correlations on a geometrically frustrated lattice causes the individual two-dimensional layers of spin- $\frac{1}{2}$ Cu^{2+} pairs (spin dimers) to become decoupled at the QCP, giving rise to a two-dimensional QCP characterized by linear power law scaling distinctly different from that of its three-dimensional counterpart. Thus the very notion of dimensionality can be said to acquire an ‘emergent’ nature: although the individual particles move on a three-dimensional lattice, their collective behaviour occurs in lower-dimensional space.

$\text{BaCuSi}_2\text{O}_6$ is a spin dimer system whose highly symmetric crystal structure^{17,18} (Fig. 1 inset) gives it unique advantages for tackling the fundamental role of dimensionality in the field of quantum criticality. The material consists of layers of spin dimers arranged vertically on an essentially square lattice in which exchange interactions are rotationally invariant around the crystalline c axis—providing the ideal conditions for Bose–Einstein condensation. However, the dimer plaquettes are staggered between consecutive layers, as shown in Fig. 1 inset. This leads to geometrical frustration of the inter-layer antiferromagnetic interaction, J_f (that is, the antiparallel arrangement of spins within and between layers conflict).

In zero magnetic field, each spin dimer has a singlet (spin $s = 0$) ground state comprising $s = \frac{1}{2}$ Cu^{2+} ions paired by an antiferromagnetic coupling constant $J = 4.45$ meV (refs 12, 13, 19), with three degenerate triplet excited levels ($s = 1$, $s_z = 1, 0, -1$). Hence the ground state of the system is a quantum paramagnet consisting of a direct product of singlets, while the single triplet excitations become dispersive owing to the moderate intra-layer coupling $J' = 0.51$ meV $\ll J$ (refs 12, 13), but remain gapped. The effect of a magnetic field is to Zeeman-split the triplet bands in a linear fashion, closing the gap between the singlet level and the lowest energy $s_z = 1$ triplet excitation (at wavevector $k_x = k_y = \pi$) at a critical magnetic field H_{c1} , which leads to an ordered state with a precisely field-tunable concentration (or density) $\rho \equiv m_z \equiv \langle s_z \rangle$ of triplets above H_{c1} (here

m_z is the uniform magnetization and $\langle s_z \rangle$ is the expectation value of s_z).

Figure 1 shows the measured phase boundary separating the quantum paramagnetic phase from the magnetically ordered state. Ordering transitions are measured by torque magnetometry¹³ in static magnetic fields of up to 28 T in the temperature range 30 mK to 1 K in a dilution refrigerator in the National High Magnetic Field Laboratory, Tallahassee. A finite torque is obtained by inclining H at a small ($<10^\circ$) angle to the crystalline c axis, from which $m_z \equiv \rho$ is extracted (see below).

We extract the critical scaling behaviour from the phase boundary in Fig. 1 to determine the universality class of the QCP. Figure 2 shows the experimental value of the critical exponent ν determined from fitting points on the phase boundary in a sliding window. The exponent ν characterizes the power law^{1,13} relating the ordering temperature to the proximity to the critical magnetic field H_{c1} : $T_c \propto (H - H_{c1})^\nu$. As the temperature is lowered, the system enters the region of universal behaviour (bright yellow shading). The value of ν tends to $\frac{2}{3}$ in this region as T tends to 0.5 K from above ($T_{\text{win}} \approx 0.9$ K), consistent with previous measurements¹³. However,

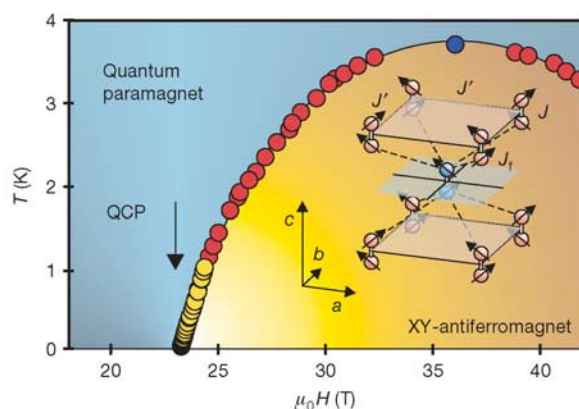


Figure 1 | Experimentally obtained phase boundary of $\text{BaCuSi}_2\text{O}_6$. The phase transition to the magnetically ordered state is shown over the entire temperature range (error bars are smaller than the symbol size); the shading varies radially from dark to light, reflecting the reduction in dimensionality near the QCP. Experimental data obtained from torque magnetometry in a dilution refrigerator are in yellow, previous experimental data from ref. 13 are in red (torque magnetometry and magnetocaloric effect in a ^3He refrigerator) and blue (specific heat in a ^4He cryostat). The inset shows a schematic diagram representing the body-centred tetragonal $\text{BaCuSi}_2\text{O}_6$ crystal lattice (the lattice exhibits a subtle incommensurate distortion below 85 K, ref. 18), dimers formed from $\text{Cu}^{2+} s = \frac{1}{2}$ spins are shown as dumb-bells. It is apparent that this lattice structure leads to geometrical frustration. J is the intra-dimer antiferromagnetic interaction, J' the antiferromagnetic interaction between in-plane dimers, and J_f the inter-layer antiferromagnetic interaction that leads to geometric frustration.

¹Geballe Laboratory for Advanced Materials and Department of Applied Physics, Stanford University, Stanford, California 94305, USA. ²NHML, MS-E536, Los Alamos National Laboratory, ³Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ⁴National High Magnetic Field Laboratory, Tallahassee, Florida 32310, USA. ⁵Institute for Solid State Physics, University of Tokyo, Kashiwa, Chiba 277-8581, Japan.

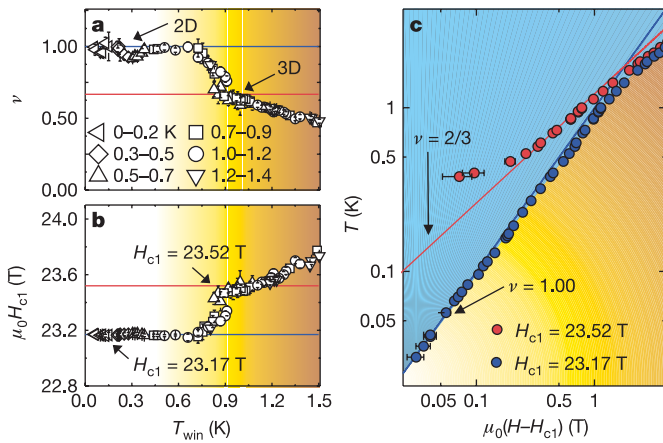


Figure 2 | Crossover from 3D to 2D BEC critical exponent. **a**, Values of the critical exponent ν obtained from fitting experimental points on the phase boundary in a sliding window centred at T_{win} (K). A two-parameter least squares regression of equation (1c) with ν and critical field H_{c1} varying is used to fit the data. Error bars show the standard errors in the least squares fit. The window size varies from 0.05 to 1.4 K, as indicated by different symbols. The data approach $\nu = \frac{2}{3}$ in the intermediate regime, and there is a distinct crossover toward $\nu = 1$ before the QCP is reached. **b**, Estimates of H_{c1} obtained along with ν during the fit. H_{c1} approaches a transient value of 23.52 T in the intermediate regime, and crosses over to the true low temperature value of 23.17 T before the QCP is reached. The shading reflects the crossover toward the $\nu = 1$ exponent as the QCP is approached both as a function of field and temperature. The dark yellow shading indicates the high temperature (field) non-universal regime, the bright yellow the intermediate regime, and the light yellow the 2D regime. **c**, Best fits to the phase boundary in the intermediate and low temperature regimes represented on a logarithmic scale. The solid lines show that the data in the intermediate temperature range are consistent with the values of $H_{c1} = 23.52$ T and $\nu = \frac{2}{3}$, which does not fit the lower temperature points; whereas data in the lower temperature range are consistent with $H_{c1} = 23.17$ T and $\nu = 1$, which does not fit the higher temperature points. We observe a crossover from one regime to the other in the temperature range $0.65 \text{ K} < T_{\text{win}} < 0.9 \text{ K}$.

below these temperatures there is a clear crossover to a value of $\nu = 1$ for the fitting range $T < 1 \text{ K}$ ($T_{\text{win}} < 0.65 \text{ K}$). We show below that these exponents are characteristic of the three- and two-dimensional Bose–Einstein condensate (3D and 2D BEC) universality classes, respectively.

The system can be considered as a 3D Bose gas of interacting particles by neglecting the unoccupied higher energy $s_z = -1, 0$ triplet states (since $J_{\text{f}}, J' \ll J$), and replacing each dimer by an effective site that can be either empty (singlet state) or occupied by a hardcore boson ($s_z = 1$ triplet state)^{12,20,21}. H acts as a chemical potential, which at low magnetic fields is prohibitively high, and prevents population of the Bose gas, permitting population only above H_{c1} . The interacting bosons move on the frustrated lattice with kinetic energy provided by the xy -component of the inter-dimer Heisenberg interaction (which determines the effective mass), while the short range repulsion arises from the Ising or z -component. The kinetic energy term dominates in $\text{BaCuSi}_2\text{O}_6$, resulting in a BEC ordered state^{12,22} characterized by a quantum-coherent superposition of the singlet and the $s_z = 1$ triplet on all sites. This can be described as a canted XY-antiferromagnet in terms of the original spin degrees of freedom (shown in Fig. 1 inset), with the in-plane staggered magnetization m_{xy} determining the amplitude and phase of the BEC order parameter $\langle b^\dagger \rangle$.

At the QCP (H_{c1}), the characteristic length of the amplitude fluctuations of $\langle b^\dagger \rangle$ diverges. Taking the scaling limit (lattice parameter $\rightarrow 0$), the original microscopic theory becomes an effective continuous field theory describing a dilute Bose gas^{1,23}. In this field theory, the dynamical critical exponent $z = 2$, and hence the

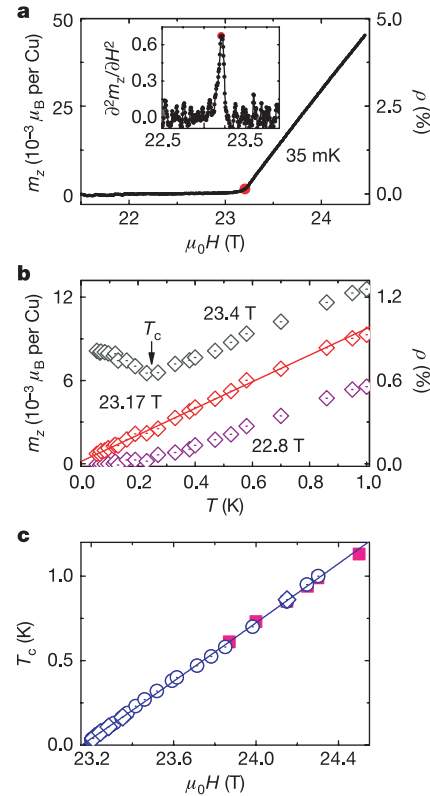


Figure 3 | 2D BEC power law behaviour. Power laws extracted from uniform magnetization obtained from torque measurements on $\text{BaCuSi}_2\text{O}_6$ in an external magnetic field applied at a small angle ($< 10^\circ$) to the crystalline c axis. Error bars are smaller than the symbol size. A small background subtraction has been made to account for the cantilever response; absolute values of magnetization were obtained by comparison with pulsed magnetic field data. **a**, Uniform magnetization ($m_z \equiv$ particle density (ρ)) obtained from torque measured as a function of rising magnetic field at 35 mK. The ordering transition determined from a sharp feature in the second derivative¹³ (shown in the inset) is indicated on the magnetization curve (**b**) $m_z \equiv \rho$ as a function of temperature extracted from the magnetization curves in **a** measured at various temperatures. Representative curves are shown at magnetic fields above, below and at the critical magnetic field $H_{c1} = 23.17 \text{ T}$. $T_c = 0 \text{ K}$ at H_{c1} , and corresponds to the dip in m_z at fields above H_{c1} . The solid line shows a linear fit to m_z at H_{c1} . **c**, Points on the phase boundary obtained from ordering transitions determined from the magnetization curves in **a**. The solid squares represent data previously reported¹³. The open symbols represent data reported in this work which are measured at lower temperatures in a dilution refrigerator. The open circles and diamonds represent data taken with the sample in slightly different orientations (the angle was changed by $\sim 10^\circ$ and the field accordingly rescaled by a 0.5% change in g -factor) to verify that the results are independent of orientation. The solid line shows a linear fit to the phase boundary below 1 K.

upper critical dimension of this QCP is $d_c = 2$. The correct critical exponents for $d \geq 2$ are therefore obtained using a mean-field theory, which leads to the following universal power laws for measurable quantities on approaching the QCP:

$$\rho(T=0) \equiv m_z(T=0) \propto (H - H_{c1}) \quad (1a)$$

$$m_z(H_{c1}) \propto T^{d/2} \quad (1b)$$

$$T_c \propto (H - H_{c1})^{2/d} \quad (1c)$$

that is, the critical exponent $\nu \equiv 2/d$ for the BEC universality class.

According to equation (1c), the experimentally measured value of $\nu = \frac{2}{3}$ as T tends to 0.5 K from above ($T_{\text{win}} \approx 0.9 \text{ K}$, Fig. 2) is consistent with a 3D BEC^{23–25}, in agreement with previous measurements¹³. In contrast, the unexpected crossover to $\nu = 1$ in the fitting range

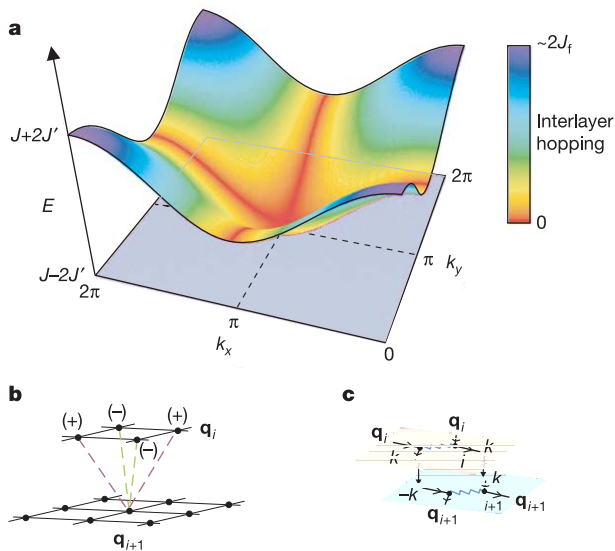


Figure 4 | Inter-layer decoupling at the QCP. **a**, Dispersing $s_z = 1$ triplon band in an applied magnetic field of H_{c1} . The shape of the dispersion is the same in any magnetic field $H \leq H_{c1}$, but the gap to the singlet state decreases with increasing H . At $H = H_{c1}$, the gap to the singlet state is closed at wavevector $k_x = k_y = \pi$, resulting in Bose–Einstein condensation. The colour shading represents the inter-layer hopping. At the wavevector where Bose–Einstein condensation occurs, there is no inter-layer hopping, which leads to a 2D BEC QCP. **b**, A diagram of the phase alternation of a single boson state with $\mathbf{q} = (\pi, \pi)$, resulting in phase cancellation on the next layer (that is, no interaction between the two layers). **c**, Inter-layer tunnelling due to zero-point fluctuations is equivalently represented as interaction-induced pair hopping, proportional to ρ^2 at low boson densities.

$T < 1$ K ($T_{\text{win}} < 0.65$ K, Fig. 2) is a signature of 2D BEC behaviour. All experimental results for $T < 1$ K in Fig. 3a–c are consistent with the linear power law behaviours predicted by equation (1) for $d = 2$, thus demonstrating that the measured critical behaviour of the system in the temperature range 30 mK to 1 K belongs to the 2D BEC universality class. Ising-like order due to U(1) symmetry breaking terms can be ruled out, as this would lead to a critical exponent of 0.5 (ref. 1). The effect of disorder would be either a critical exponent satisfying the Harris criterion^{24,26,27} ($\nu > 4/3$), or a smeared phase transition. Neither the experimentally measured exponent ($\nu = 1$), nor the sharp non-hysteretic transition in the magnetization characterized by the narrow peak in d^2m_z/dH^2 which sharpens at low temperatures, are consistent with a disordered system.

To understand the origin of the dimensional reduction, we consider the effect of geometrical frustration as a consequence of the body-centred tetragonal lattice on which the Bose gas moves. The single particle dispersion relation comprises a term arising from the intra-layer exchange, $J'(\cos k_x + \cos k_y)$, and a term reflecting the inter-layer hopping, $2J_f \cos k_z \cos(k_x/2) \cos(k_y/2)$, as shown in Fig. 4a. At the dispersion minimum ($k_x = k_y = \pi$) where the bosons condense, the inter-layer hopping is always zero (that is, there is no dispersion along the c axis). Equivalently, the phase of a single boson (with $k_x = k_y = \pi$) alternates cyclically on a plaquette of neighbouring lattice sites (shown in Fig. 4b), resulting in phase cancellation at each site of the adjacent layer. Hence the effect of geometrical frustration is to decouple adjacent layers, leading to a highly degenerate ground state, which can be described as an array of 2D BEC layers.

A competing effect arises from zero-point phase fluctuations, which generate an effective inter-layer coupling (K) and restore phase coherence along the c axis^{28,29}. However, because K can equivalently be considered to be a consequence of pair tunnelling (as shown in Fig. 4c) and hence biquadratic in the order parameter, its strength decreases as ρ^2 for low densities of bosons (as recently

shown using a spin-wave approach on the same spin lattice²⁸). A measure of the ‘3D-ness’ of the system is given by the size of the effective coupling K relative to the temperature scale $k_B T_c$. As T_c is proportional to ρ in the proximity of the QCP (from equation (1) for $d = 2$), the ratio $K/k_B T_c$ (which is $\propto \rho$) is arbitrarily small close enough to the QCP, and inter-layer tunnelling is too small for the system to be observably 3D. By field tuning toward H_{c1} , we experimentally access this region, and hence observe critical power law behaviour consistent with the 2D BEC universality class. The energy scale of this 2D-critical behaviour is well separated from the very low temperatures at which weak longer range inter-layer interactions are anticipated to restore 3D behaviour; and hence 2D behaviour is observed over a significant range of temperature.

Away from the QCP, the fluctuation-induced inter-layer tunnelling increases rapidly with particle density, leading to distinctly 3D behaviour when $K(\rho)$ becomes comparable to T_c . The system may then be described by an effective unfrustrated model for the subsets of odd and even layers that assumes a constant effective inter-layer coupling, J'' (refs 12, 13). Tuning the system by means of the applied magnetic field from this unfrustrated region toward the QCP where geometrical frustration becomes effective therefore results in a marked crossover from 3D to 2D BEC power law behaviour (Fig. 2).

BaCuSi₂O₆ therefore provides a clear example of a system in which geometrical frustration causes the effective dimensionality to become reduced at the QCP, leading to 2D collective excitations despite the 3D nature of the system. Although inter-layer decoupling due to geometrical frustration features as a possible explanation for the puzzling experimental observation of non-Fermi-liquid behaviour at the QCP in body-centred tetragonal heavy fermion intermetallics^{3–8}, theoretical reasoning²⁸ has been used to assert that it is impossible to realize reduced dimensionality in any system by the mechanism of geometrical frustration. The experimentally observed dimensional reduction in BaCuSi₂O₆ provides a counter-example, and constitutes a proof of principle that dimensionality can become an emergent property of a QCP.

Received 25 August 2005; accepted 13 March 2006.

- Sachdev, S. *Quantum Phase Transitions* (Cambridge Univ. Press, New York, 1999).
- Hertz, J. Quantum critical phenomena. *Phys. Rev. B* **14**, 1165–1184 (1976).
- Sachdev, S. Quantum criticality: Competing ground states in low dimensions. *Science* **288**, 475–480 (2000).
- Stockert, O., von Lohneysen, H., Rosch, A., Pyka, N. & Loewenhaupt, M. Two-dimensional fluctuations at the quantum-critical point of CeCu_{6-x}Au_x. *Phys. Rev. Lett.* **80**, 5627–5630 (1998).
- Si, Q., Rabello, S., Ingersent, K. & Smith, J. L. Locally critical quantum phase transitions in strongly correlated metals. *Nature* **413**, 804–808 (2001).
- Mathur, N. D. et al. Magnetically mediated superconductivity in heavy fermion compounds. *Nature* **394**, 39–43 (1998).
- Senthil, T., Vishwanath, A., Balents, L., Sachdev, S. & Fisher, M. P. A. Deconfined quantum critical points. *Science* **303**, 1490–1494 (2004).
- Millis, A. J. Effect of a nonzero temperature on quantum critical points in itinerant fermion systems. *Phys. Rev. B* **48**, 7183–7196 (1993).
- Xu, C. & Moore, J. E. Geometric criticality for transitions between plaquette phases in integer-spin Kagome XXZ antiferromagnets. *Phys. Rev. B* **72**, 064455 (2005).
- Batista, C. D. & Nussinov, Z. Generalized Elitzur’s theorem and dimensional reductions. *Phys. Rev. B* **72**, 045137 (2005).
- Tewari, S., Toner, J. & Chakravarty, S. Floating phase in a dissipative Josephson junction array. *Phys. Rev. B* **72**, 060505(R) (2005).
- Jaime, M. et al. Magnetic-field-induced condensation of triplons in Han purple pigment BaCuSi₂O₆. *Phys. Rev. Lett.* **93**, 087203 (2004).
- Sebastian, S. E. et al. Characteristic Bose–Einstein condensation scaling close to a quantum critical point in BaCuSi₂O₆. *Phys. Rev. B* **72**, 100404(R) (2005).
- FitzHugh, E. W. & Zycherman, L. A. A purple barium copper silicate pigment from early China. *Studies Conserv.* **37**, 145–154 (1992).
- Nikuni, T., Oshikawa, M., Oosawa, A. & Tanaka, H. Bose–Einstein condensation of diluted magnons in TiCuCl₃. *Phys. Rev. Lett.* **84**, 5868–5871 (2000).
- Ruegg, Ch. et al. Bose–Einstein condensation of the triplet states in the magnetic insulator TiCuCl₃. *Nature* **423**, 62–65 (2003).
- Sparta, K. M. & Roth, G. Reinvestigation of the structure of BaCuSi₂O₆—evidence for a phase transition at high temperature. *Acta Crystallogr. B* **60**, 491–495 (2004).

18. Samulon, E. C. *et al.* Low temperature structural phase transition and incommensurate lattice modulation in the spin gap compound $\text{BaCuSi}_2\text{O}_6$. *Phys. Rev. B* **73**, 100407(R) (2006).
19. Sasago, Y., Uchinokura, K., Zheludev, A. & Shirane, A. Temperature-dependent spin gap and singlet ground state in $\text{BaCuSi}_2\text{O}_6$. *Phys. Rev. B* **55**, 8357–8360 (1991).
20. Mila, F. Ladders in a magnetic field: a strong coupling approach. *Eur. Phys. J. B* **6**, 201–205 (1998).
21. Giamarchi, T. & Tsvelik, A. M. Coupled ladders in a magnetic field. *Phys. Rev. B* **59**, 11398–11407 (1999).
22. Rice, T. M. To condense or not to condense. *Science* **298**, 760–761 (2002).
23. Kawashima, N. Quantum critical point of the XY model and condensation of field-induced quasiparticles in dimer compounds. *J. Phys. Soc. Jpn* **73**, 3219–3222 (2004).
24. Fisher, M. P. A. *et al.* Boson localization and the superfluid-insulator transition. *Phys. Rev. B* **40**, 546–570 (1989).
25. Nohadani, O. *et al.* Universal scaling at field-induced magnetic phase transitions. *Phys. Rev. B* **69**, 220402(R) (2004).
26. Harris, A. B. Effect of random defects on the critical behaviour of Ising models. *J. Phys. C* **7**, 1671–1692 (1974).
27. Vojta, T. & Sknepnek, R. Critical points and quenched disorder: from Harris criterion to rare regions and smearing. *Phys. Status Solidi B* **241**, 2118–2127 (2004).
28. Maltseva, M. & Coleman, P. Failure of geometric frustration to preserve a quasi-two-dimensional spin fluid. *Phys. Rev. B* **72**, 174415(R) (2005).
29. Henley, C. L. Ordering due to disorder in a frustrated vector antiferromagnet. *Phys. Rev. Lett.* **62**, 2056–2059 (1989).

Acknowledgements N.H., C.D.B., M.J. and P.A.S. acknowledge Laboratory Directed Research and Development (LDRD) support at LANL. S.E.S. and I.R.F. acknowledge National Science Foundation (NSF) support. Experiments performed at the NHMFL, Tallahassee, were supported by the NSF, the State of Florida, and the Department of Energy. We thank T. P. Murphy, E. C. Palm, P. Tanedo and P. B. Brooks for experimental assistance, and acknowledge discussions with A. G. Green, E.-A. Kim, S. A. Kivelson, D. I. Santiago and J. Schmalian. I.R.F. acknowledges support from the Alfred P. Sloan Foundation and S.E.S. from the Mustard Seed Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.E.S. (suchitra@stanfordalumni.org).

Thermodynamic control of asymmetric amplification in amino acid catalysis

Martin Klussmann¹, Hiroshi Iwamura^{1,†}, Suju P. Mathew¹, David H. Wells Jr^{1,†}, Urvish Pandya¹, Alan Armstrong¹ & Donna G. Blackmond^{1,2}

Ever since Pasteur noticed that tartrate crystals exist in two non-superimposable forms that are mirror images of one another—as are left and right hands—the phenomenon of chirality has intrigued scientists. On the molecular level, chirality often has a profound impact on recognition and interaction events and is thus important to biochemistry and pharmacology. In chemical synthesis, much effort has been directed towards developing asymmetric synthesis strategies that yield product molecules with a significant excess of either the left-handed or right-handed enantiomer. This is usually achieved by making use of chiral auxiliaries or catalysts that influence the course of a reaction, with the enantiomeric excess (ee) of the product linearly related to the ee of the auxiliary or catalyst used. In recent years, however, an increasing number of asymmetric reactions have been documented where this relationship is nonlinear¹, an effect that can lead to asymmetric amplification. Theoretical models^{2,3} have long suggested that autocatalytic processes can result in kinetically controlled asymmetric amplification, a prediction that has now been verified experimentally^{4–6} and rationalized mechanistically^{7–14} for an autocatalytic alkylation reaction. Here we show an alternative mechanism that gives rise to asymmetric amplification based on the equilibrium solid-liquid phase behaviour of amino acids in solution. This amplification mechanism is robust and can operate in aqueous systems, making it an appealing proposition for explaining one of the most tantalizing examples of asymmetric amplification—the development of high enantiomeric excess in biomolecules from a presumably racemic prebiotic world.

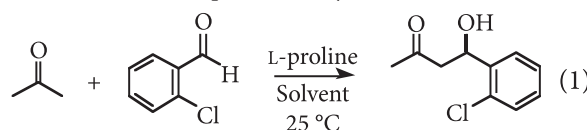
Recent investigations have rekindled interest in exploring nonlinear effects in amino acid catalysis^{15,16} despite experimental and theoretical work¹⁷ that has effectively discounted earlier reports¹⁸ of nonlinear behaviour in these systems. Our measurements on the proline-catalysed aldol reaction (1) (see Supplementary Information for experimental details) reveal behaviour more complex than was found in either the earlier or later work: two regimes exist with different relationships between the ee of the proline catalyst and the ee of the aldol product. If the reaction is carried out using proline concentrations low enough that the catalyst is fully dissolved, the relationship is linear (Fig. 1a, squares). When using higher catalyst concentrations, so that dissolved proline is in equilibrium with solid proline, the product ee is largely independent of proline ee (Fig. 1a, circles). This latter, unusual relationship cannot be explained by any of the existing models for nonlinear effects in asymmetric catalysis¹⁹. However, this phenomenon mirrors the trend in the ee of dissolved proline as a function of total proline ee (Fig. 1b), with the proline solution ee found to remain constant at about 50% ee even though the total proline ee is varied from near racemic to near enantiopure.

The solid-solution phase behaviour of enantiomers and their

racemates was the focus of extensive study at the turn of the last century²⁰ and is generally well-understood²¹, but non-enantiopure, non-racemic mixtures (particularly those of free amino acids) have received less attention. We therefore explored the phase behaviour of proline and found that when the two enantiomers (L and D) of proline are present in unequal proportions above their solubility limit in dimethyl sulphoxide (DMSO), two separate solid phases are formed at equilibrium. Fourier-transform infrared (FTIR) spectroscopic analysis of the solids confirmed that these are a racemic compound (that is, crystals in which L:D = 1:1), and an enantiopure solid of the excess enantiomer. The ternary phase diagram for the system D-proline/L-proline/DMSO constructed from our data is shown in Fig. 2, where pure DMSO is represented at the apex, and pure D and L proline at the left and right vertices, respectively.

The low solubility of proline means that the part of the diagram that is of interest for our solution phase system is located very near to the apex, and this portion is expanded in the main part of Fig. 2. The diagram is symmetric: racemic mixtures are represented along the vertical line intersecting the DMSO vertex, and the phase behaviour of D-proline to the left side of this line reflects the L-proline phase behaviour on the right. The phase rule dictates that at constant temperature the composition of a solution of proline in equilibrium with the two solid phases is fixed, and this composition is given in Fig. 2 by the eutectic at E' (or E). Only proline compositions with very low ee (where the excess of the major enantiomer is too low to establish its solid phase) or very high ee (where the minor enantiomer concentration is insufficient to form the solid racemic compound) will show a variation of solution ee with overall proline ee, dictated by the lines E'R (or ER) and E'A' (or EA), respectively, in Fig. 2.

For proline in DMSO at 25 °C, the eutectic composition corresponds to a solution enantiomeric excess of about 50% ee. Any saturated solution of scalemic proline in equilibrium with a sufficient excess of solid will exhibit this ee, regardless of the overall ee of the proline employed. Given that the proline-catalysed aldol reaction (1) occurs in the solution phase, this finding readily explains the relationships between proline ee and the product ee illustrated in Fig. 1a; that is, the ee of the aldol product appears to depend linearly on the ee of the accessible proline catalyst.



Phase diagrams such as the one shown in Fig. 2 are useful for optimizing crystallization conditions, where the aim is typically the production of a solid of high enantiopurity by depleting the solution

¹Department of Chemistry and ²Department of Chemical Engineering and Chemical Technology, Imperial College, London SW7 2AZ, UK. †Present addresses: Mitsubishi Pharma, 14, Sunayama, Kamisu, Ibaraki 314-0255, Japan (H.I.); University of Calgary, 2500 University Drive NW, Calgary AB, T2N 1N4, Canada (D.H.W.).

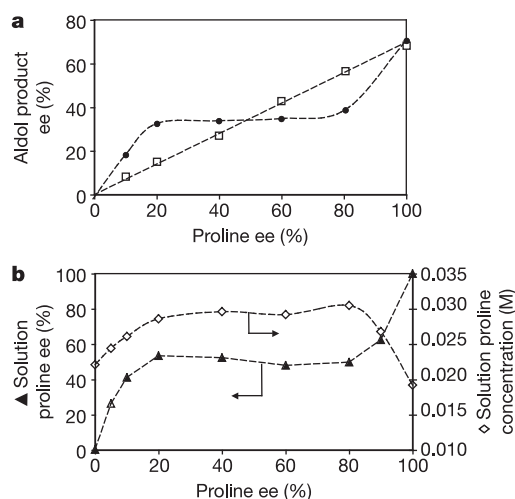


Figure 1 | Reaction and solution behaviour as a function of the overall proline enantiomeric excess. **a**, Product enantiomeric excess versus proline enantiomeric excess for the aldol reaction of equation (1). Squares, proline concentration 0.025 M; circles, total proline concentration 0.1 M (corresponding to ~4 mol% solution proline). **b**, Solution proline enantiomeric excess (left axis, triangles) and solution proline concentration (right axis, diamonds) as a function of the overall enantiomeric excess for proline at 0.1 M; all experiments in DMSO with 0.8 wt% H₂O at 25 °C.

enantiomeric excess. However, our findings suggest that aiming for the opposite effect—high solution enantiomeric excess and lower solid ee—provides a means of realizing asymmetric amplification in solution. This prompted us to examine the phase behaviour of a number of amino acids, to explore their potential for higher asymmetric amplification than is possible with proline and its eutectic positioned at about 50% ee. Although all but two of the twenty proteinogenic amino acids are known to form racemic compounds (crystals with a 1:1 ratio of the D and L enantiomers) and not conglomerates (a mixture of pure D and pure L crystals), we were unable to find literature values for the eutectic composition of any ternary phase system of free amino acids that form racemic

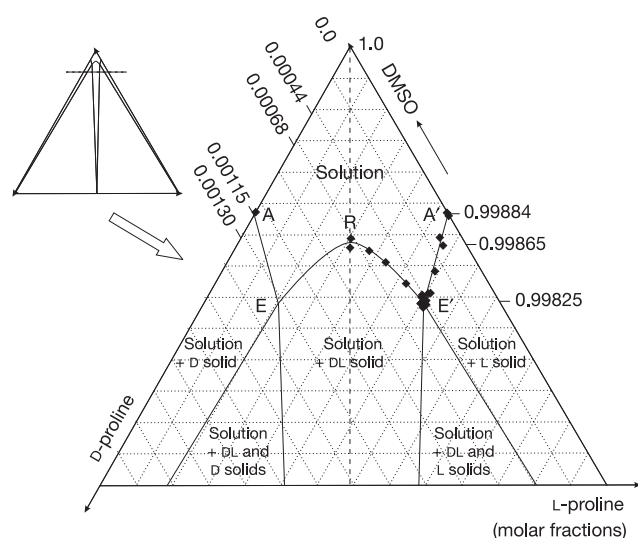


Figure 2 | Ternary phase diagram of D-proline, L-proline and DMSO at 25 °C. The main figure expands the area above the dashed line in the inset (not to scale). Saturated solutions exist along the line A-E and A-E'. E and E' represent the eutectic mixture composition where three phases (solid enantiopure proline, solid racemic compound, and liquid phase) exist in equilibrium.

Table 1 | Solution enantiomeric excess at the eutectic point in water at 25 °C for selected amino acids

Amino acid	ee of solution at eutectic (%)	Amino acid	ee of solution at eutectic (%)
Threonine	0	Methionine	85
Valine	46	Leucine	87
Alanine	60	Histidine	93
Phenylalanine	83	Serine	>99

compounds in water or other solvent. The two amino acids that form conglomerates, threonine²² and arginine, exhibit 0% ee at the eutectic. Our own measurements collated in Table 1 reveal that several of the proteinogenic amino acids that form racemic compounds in fact exhibit high eutectic ee values. Serine, with its eutectic at >99% ee, provides a virtually enantiopure solution from a nearly racemic sample under solid-liquid equilibrium conditions.

Prediction of enantioselectivity in solution catalysis from knowledge of these eutectics was confirmed in the results of the aldol reaction of equation (1) carried out using instead of L-proline several of the amino acids listed in Table 1 at varying levels of enantiopurity, with three-phase equilibrium established before reaction. Figure 3 shows how the product enantiomeric excesses obtained with catalysts of varying ee values compare with the product ee that is obtained when using enantiopure amino acid catalyst. The data indicate that the position of the eutectic dictates the product selectivity for this amino-acid-catalysed transformation. Product enantioselectivities for aldol reactions carried out in aqueous media using these amino acid catalysts were also found to correlate well with the position of the eutectic, although yields and chemo- and enantioselectivities were in general lower (see Supplementary Information). As expected from the eutectic values listed in Table 1, serine offers the most significant chiral amplification: nearly racemic (1% ee) serine gives a product enantioselectivity virtually indistinguishable from that obtained using enantiopure serine.

Here we have established that the coupling of solid-liquid phase behaviour and amino acid catalysis results in an efficient and robust mechanism for asymmetric amplification. This finding suggests that amino acids may have played a role in the evolution of biomolecular

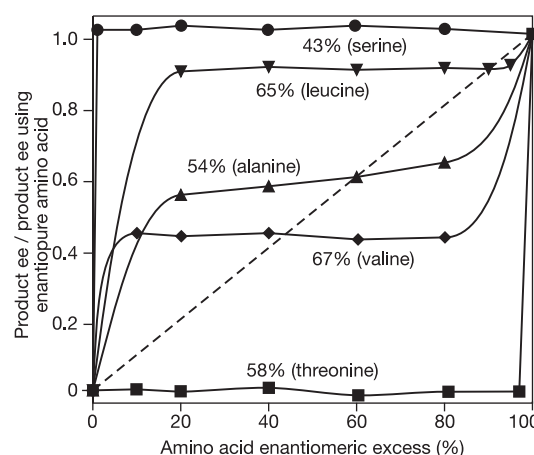


Figure 3 | Nonlinear effects in an amino-acid mediated aldol reaction. Product enantiomeric excess expressed as a fraction of that obtained using enantiopure amino acid versus amino acid enantiomeric excess in the aldol reaction of equation (1) carried out in N,N-dimethylformamide (DMF) at ambient temperature using serine (circles), leucine (down triangles), alanine (up triangles), valine (diamonds), and threonine (squares). Percentage product ee values found for the enantiopure catalysts are given.

homochirality²³. One can easily imagine the coexistence of solid and dissolved amino acids under a range of environmental conditions in a prebiotic landscape, and high ee could then evolve from an initially small imbalance in ee if (1) the amino acid forms a racemic compound rather than a conglomerate; (2) the amino acid exhibits high enantiopurity at its eutectic; and (3) the amino acid acts with high selectivity as an asymmetric catalyst in solution phase transformations. In fact, our studies show that serine meets the first two criteria, and recent work²⁴ has demonstrated that serine catalyses certain asymmetric aldol reactions with high enantioselectivity; it has also been highlighted as one of a small number of amino acids likely to have played an important role in prebiotic chemistry^{25,26}.

A scenario for homochiral evolution based on our amplification mechanism was hinted at some years ago²⁷, and it also shares some similarities with the suggestion that chiral enrichment may result from the selective dissolution of only one (enantiopure) crystal of a conglomerate in a microenvironment²⁸, but seems ultimately much simpler. A particularly appealing feature of the scenario we outline here is that it is based on an equilibrium mechanism, in contrast to the far-from-equilibrium environments invoked in kinetically induced amplification via autocatalytic reactions or crystallizations^{29,30}. Finally, we note that further enhancements in solution-phase enantiomeric excess may be discovered in more complex multi-component systems with a greater number of equilibrium solid and liquid phases. This possibility and the influence of interactions between different amino acids and between amino acids and other molecules such as sugars is the subject of ongoing investigations by our group, as is the search for high selectivity in amino-acid-catalysed reactions of potential biological relevance.

Received 21 November 2005; accepted 30 March 2006.

- Girard, C. & Kagan, H. B. Nonlinear effects in asymmetric synthesis and stereoselective reactions: ten years of investigation. *Angew. Chem. Int. Edn* **37**, 2923–2959 (1998).
- Frank, F. C. Spontaneous asymmetric synthesis. *Biochim. Biophys. Acta* **11**, 459–463 (1953).
- Calvin, M. *Molecular Evolution* (Oxford Univ. Press, Oxford, UK, 1969).
- Shibata, T., Morioka, H., Hayase, T., Choji, K. & Soai, K. Highly enantioselective catalytic asymmetric automultiplication of chiral pyrimidyl alcohol. *J. Am. Chem. Soc.* **118**, 471–472 (1996).
- Shibata, T., Choji, K., Hayase, T., Aizu, Y. & Soai, K. Asymmetric autocatalytic reaction of 3-quinolylalkanol with amplification of enantiomeric excess. *Chem. Commun.*, 1235–1236 (1996).
- Soai, K., Shibata, T., Morioka, H. & Choji, K. Asymmetric autocatalysis and amplification of enantiomeric excess of a chiral molecule. *Nature* **378**, 767–768 (1995).
- Blackmond, D. G., McMillan, C. R., Ramdeehul, S., Schorm, A. & Brown, J. M. Origins of asymmetric amplification in autocatalytic alkylzinc additions. *J. Am. Chem. Soc.* **123**, 10103–10104 (2001).
- Blackmond, D. G. Description of the condition for asymmetric amplification in autocatalytic reactions. *Adv. Synth. Catal.* **344**, 156–158 (2002).
- Buono, F. G. & Blackmond, D. G. Kinetic evidence for a tetrameric transition state in the asymmetric autocatalytic alkylation of pyrimidyl aldehydes. *J. Am. Chem. Soc.* **125**, 8978–8979 (2003).
- Buono, F. G., Iwamura, H. & Blackmond, D. G. Physical and chemical rationalization for asymmetric amplification in autocatalytic reactions. *Angew. Chem. Int. Edn* **43**, 2099–2103 (2004).
- Blackmond, D. G. Asymmetric autocatalysis and its implications for the origin of homochirality. *Proc. Natl. Acad. Sci. USA* **101**, 5732–5736 (2004).
- Gridnev, I. D. & Brown, J. M. Asymmetric autocatalysis: novel structures, novel mechanism? *Proc. Natl. Acad. Sci. USA* **101**, 5727–5731 (2004).
- Gridnev, I. D., Serafimov, J. M. & Brown, J. M. Solution structure and reagent binding of the zinc alkoxide catalyst in the Soai asymmetric autocatalytic reaction. *Angew. Chem. Int. Edn* **43**, 4884–4887 (2004).
- Gridnev, I. D., Serafimov, J. M., Quiney, H. & Brown, J. M. Reflections on spontaneous asymmetric synthesis by amplifying autocatalysis. *Org. Biomol. Chem.* **1**, 3811–3819 (2003).
- Mathew, S. P., Iwamura, H. & Blackmond, D. G. Amplification of enantiomeric excess in a proline-mediated reaction. *Angew. Chem. Int. Edn* **43**, 3317–3321 (2004).
- Iwamura, H., Mathew, S. P. & Blackmond, D. G. In situ catalyst improvement in the proline-mediated α -amination of aldehydes. *J. Am. Chem. Soc.* **126**, 11770–11771 (2004).
- Hoang, L., Bahmanyar, S., Houk, K. N. & List, B. Kinetic and stereochemical evidence for the involvement of only one proline molecule in the transition states of proline-catalyzed intra- and intermolecular aldol reactions. *J. Am. Chem. Soc.* **125**, 16–17 (2003).
- Agami, C. Mechanism of the proline-catalyzed enantioselective aldol reaction. Recent advances. *Bull. Soc. Chim. Fr.* (3), 499–507 (1988).
- Blackmond, D. G. Kinetic aspects of nonlinear effects in asymmetric catalysis. *Acc. Chem. Res.* **33**, 402–411 (2000).
- Roozeboom, H. W. B. Solubility and melting-point as criteria for racemate compounds, pseudoracemic mix-crystals and inactive conglomerates. *Z. Phys. Chem. Stoechiometrie Verwandtschaftslehre* **28**, 494–517 (1899).
- Jacques, J., Collet, A. & Wilen, S. H. *Enantiomers, Racemates and Resolution* Ch. 3 (John Wiley, New York, 1981).
- Rodrigo, A. A., Lorenz, H. & Seidel-Morgenstern, A. Online monitoring of preferential crystallization of enantiomers. *Chirality* **16**, 499–508 (2004).
- Podlech, J. Origin of organic molecules and biomolecular homochirality. *Cell. Mol. Life Sci.* **58**, 44–60 (2001).
- Cordova, A. et al. Acyclic amino acid-catalyzed direct asymmetric aldol reactions: alanine, the simplest stereoselective organocatalyst. *Chem. Commun.* 3586 (2005).
- Borsenberger, V. et al. Exploratory studies to investigate a linked prebiotic origin of RNA and coded peptides. *Chem. Biodivers.* **1**, 203–246 (2004).
- Takats, Z., Nanita, S. C. & Cooks, R. G. Serine octamer reactions: indicators of prebiotic relevance. *Angew. Chem. Int. Edn* **42**, 3521–3523 (2003).
- Morowitz, H. J. A mechanism for the amplification of fluctuations in racemic mixtures. *J. Theor. Biol.* **25**, 491 (1969).
- Welch, C. J. Formation of highly enantioenriched microenvironments by stochastic sorting of conglomerate crystals: a plausible mechanism for generation of enantioenrichment on the prebiotic earth. *Chirality* **13**, 425–427 (2001).
- Kondepudi, D. K. & Asakura, K. Chiral autocatalysis, spontaneous symmetry breaking, and stochastic behavior. *Acc. Chem. Res.* **34**, 946–954 (2001).
- Kondepudi, D. K., Kaufman, R. J. & Singh, N. Chiral symmetry breaking in sodium chlorate crystallization. *Science* **250**, 975–977 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding from the EPSRC and AstraZeneca is gratefully acknowledged.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.G.B. (d.blackmond@imperial.ac.uk).

LETTERS

Early stone technology on Flores and its implications for *Homo floresiensis*

Adam Brumm¹, Fachroel Aziz², Gert D. van den Bergh³, Michael J. Morwood⁴, Mark W. Moore⁴, Iwan Kurniawan², Douglas R. Hobbs⁵ & Richard Fullagar⁶

In the Soa Basin of central Flores, eastern Indonesia, stratified archaeological sites, including Mata Menge, Boa Lesa and Kobatuwa (Fig. 1), contain stone artefacts associated with the fossilized remains of *Stegodon florensis*, Komodo dragon, rat and various other taxa. These sites have been dated to 840–700 kyr BP (thousand years before present)¹. The authenticity of the Soa Basin artefacts and their provenance have been demonstrated by previous work^{2–6}, but to quell lingering doubts⁷, here we describe the context, attributes and production modes of 507 artefacts excavated at Mata Menge. We also note specific similarities, and apparent technological continuity, between the Mata Menge stone artefacts and those excavated from Late Pleistocene levels at Liang Bua cave, 50 km to the west. The latter artefacts, dated to between 95–74 and 12 kyr ago^{8,9}, are associated with the remains of a dwarfed descendent of *S. florensis*, Komodo dragon, rat and a small-bodied hominin species, *Homo floresiensis*, which had a brain size of about 400 cubic centimetres^{10,11}. The Mata Menge evidence negates claims that stone artefacts associated with *H. floresiensis* are so complex that they must have been made by modern humans (*Homo sapiens*)⁷.

Mata Menge occurs in the 80–120-m-thick Ola Bula Formation, which was deposited between 960 and 700 kyr ago in a fluvio-lacustrine environment, probably by one or more distal volcanic fans entering a lake or small series of lakes¹. The artefact- and bone-bearing deposits at Mata Menge lie between two tuffaceous silts with ascending zircon fission-track ages of 880 ± 70 kyr and 800 ± 70 kyr BP, respectively^{1,4}. The nearby site of Boa Lesa also yielded a fission-track age of 840 ± 70 kyr BP for tuffaceous silt sealing in sandstones containing stone artefacts and *Stegodon* fossils^{5,12} (see Supplementary Information). This provides a minimum age for the presence of hominins on Flores.

Our excavations at Mata Menge in 2004–05 (Fig. 2) yielded a total of 487 *in situ* stone artefacts. They had been subject to hydraulic transportation and size sorting—artefacts measuring less than 10 mm in maximum dimension appear to have been ‘winnowed’ by fluvial flow or wave action¹³. Nonetheless, the minimal degree of edge damage and abrasion to the surface indicate that the artefacts had not been exposed to prolonged water movement (52% fresh/unabraded, 35% slightly abraded and 13% heavily abraded)¹⁴. This analysis is based on the 507 stone artefacts excavated from Mata Menge in 1994 (refs 6, 12) and 2004–05 (Table 1), which comprise the largest and most securely dated Early Pleistocene assemblage from Island Southeast Asia.

Stone artefact technology at Mata Menge was simple, based on the removal of small- to medium-sized flakes from cobbles and flake

blanks. Flakes were struck invasively and non-invasively across core faces by direct freehand percussion using a cobble or recycled core as a hammerstone. Other reduction techniques included the removal of elongated flakes down the edges of cores and flakes (‘burination’) (Fig. 3f) and anvil-supported percussion (truncation) (Fig. 3d; see also Supplementary Information). Dominant raw materials were moderate to low quality, volcanic/metavolcanic, fluvial cobbles (average maximum dimension: 33.1 ± 14.5 mm; range 15.6–87.2 mm) that were obtained locally and provided raw material for 91% of the artefacts ($N = 459$). In addition, 27 fine-grained chert, 13 chalcedony, four ‘chlorite’ and four ‘opal’ artefacts were recorded.

There are five core types in the assemblage: (1) radial cores with flakes struck towards the centre of the core face from a platform edge extending partially or completely around the periphery (Fig. 3a–d, f); (2) single platform cores with flakes removed unidirectionally from a single platform edge; (3) multiplatform cores with flakes removed irregularly from different platform edges; (4) assayed cobbles with one or more amorphous scars interpreted as the result of hominins testing raw materials for flaking quality; and (5) undiagnostic cores with ambiguous flake scar patterns. An average of six flakes were removed from the cores before discard, with the maximum recorded being 26.

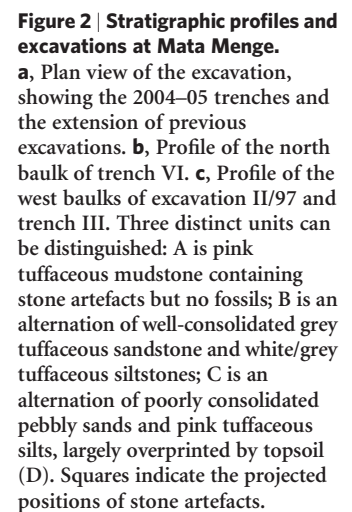
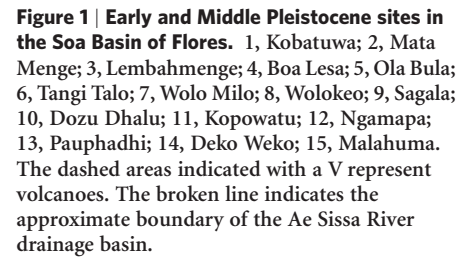
Radial core reduction of both cobbles and flake blanks or fragments was the most common flaking strategy. Some radial cores were unifacially flaked ($N = 5$) but most were worked bifacially ($N = 36$); the rest ($N = 5$) are unclear. Mata Menge knappers generally proceeded in a series of unifacial removals followed by the removal of flakes from the opposite core face, while further reduction yielded multiplatform cores.

Flakes struck from cores were generally small to medium in size (average maximum dimension: 32.4 ± 12 mm; range 5–70.7 mm) and amorphous in shape (Fig. 3g–i). Most were probably produced and discarded on the spot, but large flakes made from fine-grained materials, such as chert (for example, Fig. 3a), appear to have been carried to Mata Menge from cores struck elsewhere.

Most flakes were discarded without further modification but some exhibited microwear (Fig. 3h, i; see also Supplementary Information) or retouch, indicating use as tools. Unifacial or bifacial retouch by hard-hammer percussion was identified on 41 volcanic/metavolcanic, two chalcedony and two chert flakes (Fig. 3e). The average maximum dimension of retouched flakes (42.3 ± 12.9 mm; range 21.2–89.4 mm) was significantly larger than that of the scars on the cores (22.8 ± 10.4 mm; range 5.7–72 mm), indicating that the Mata Menge hominins generally selected larger flakes for retouching.

Most of the retouched flakes have unpatterned scars, interpreted as

¹Department of Archaeology and Natural History, Research School of Pacific and Asian Studies, Australian National University, Canberra, Australian Capital Territory 0200, Australia. ²Geological Research and Development Centre, Jalan Diponegoro 57, Bandung 40122, Indonesia. ³Naturalis, the National Museum of Natural History, P.O. Box 9517, 2300 RA Leiden, The Netherlands. ⁴Department of Archaeology and Palaeoanthropology, School of Human and Environmental Studies, University of New England, Armidale, New South Wales 2351, Australia. ⁵School of Earth and Environmental Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia. ⁶Department of Archaeology, A14, University of Sydney, Sydney, New South Wales 2006, Australia.



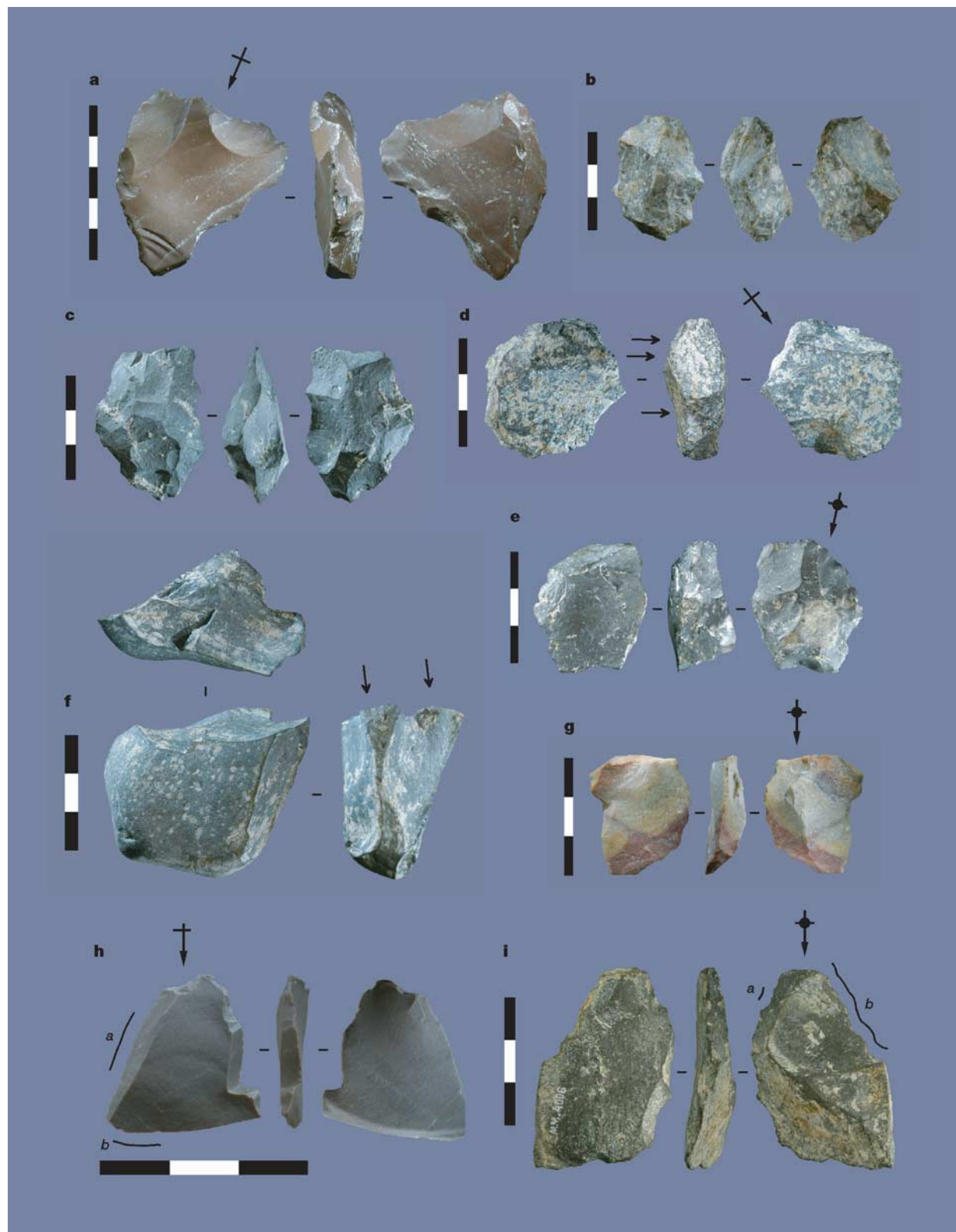


Figure 3 | Stone artefacts from Mata Menge. **a, b**, Chert bifacial radial cores; **a** is made on a flake. **c, d**, Volcanic/metavolcanic bifacial radial cores; **d** is made on a flake and features three truncation scars (indicated by small arrows). **e**, Volcanic/metavolcanic retouched flake. **f**, Volcanic/metavolcanic cobble radial core with two 'burination' scars (indicated by arrows). **g**, Chert

flake. **h**, Chert flake with microwear in the form of edge rounding (**a**) and edge scarring (**b**). **i**, Volcanic/metavolcanic flake with microwear including abrasive smoothing and a grainy polish (**a**) and small scars and striations (**b**). Scale bars are in 10-mm increments.

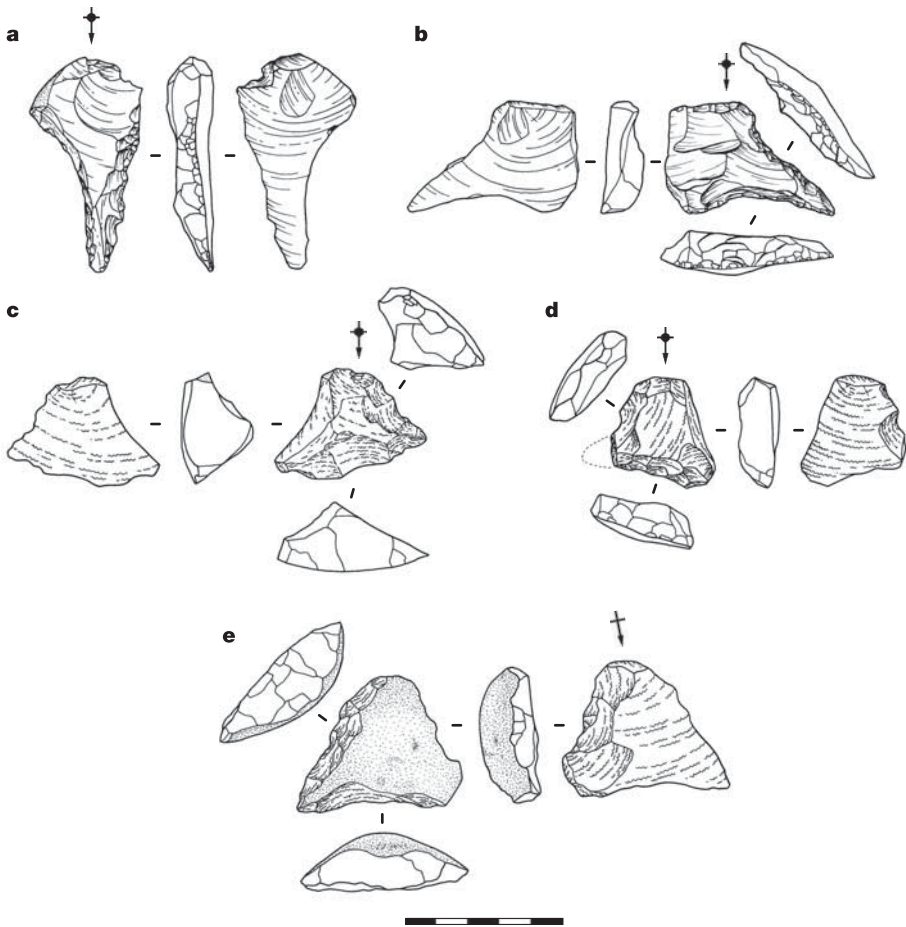


Figure 4 | Perforators from Mata Menge and Liang Bua. **a**, Chert perforator from between spits 60 and 71, sector VII, Liang Bua. **b**, Chert perforator from sector IV, spit 55, Liang Bua. **c–e**, Volcanic/metavolcanic perforators from stratified Early Pleistocene sediments at Mata Menge. Scale bar, 50 mm.

the result of casual rejuvenation of used edges, but five have extensive unifacial retouch forming a pointed projection on the margin or distal end (Fig. 4c–e). Similar deliberately shaped ‘perforators’ were also excavated at Boa Lesa¹⁵. Despite being separated by 50 km and at least 700,000 yr, there are remarkable similarities between the stone artefact assemblage from Mata Menge and that found with *H. floresiensis* at Liang Bua¹⁶ (Table 1). For instance, both assemblages show an emphasis on the use of volcanic/metavolcanic fluvial cobbles as raw materials, along with the transportation of flake blanks for use as cores. Core reduction strategies at Mata Menge and Liang Bua are also very

similar, with special emphasis on freehand reduction of cores both bifacially and radially. In fact, small, invasively reduced radial cores from the two sites are virtually indistinguishable. In addition, single platform cores, multiplatform cores, cores with ‘burination’ scars from the production of elongated flakes, ‘truncated’ flakes and cores indicating anvil-supported percussion and ‘perforators’ occur in both assemblages (Fig. 4). The maximum dimensions of flake scars on Mata Menge and Liang Bua cores are also very similar (Fig. 5; see also Supplementary Information). This is notable given that Liang Bua cores were more often on flakes, whereas Mata Menge cores were predominantly cobbles and hence tend to be bigger.

Differences between the assemblages include the fact that heat-fractured artefacts only occur at Liang Bua, whereas only one possible example of bipolar reduction of flake edges (relatively common at Liang Bua) was recorded at Mata Menge. Overall—and in spite of

Table 1 | Artefact types and frequencies from Mata Menge and Liang Bua

Type	Liang Bua		Mata Menge	
	N	Percentage	N	Percentage
Flake	3,264	90	294	57.9
Retouched flake	177	4.8	46	9.1
Radial core	103	2.8	46	9.1
Single platform core	2	0.1	7	1.3
Multiplatform core	25	0.6	10	1.9
Undiagnostic core	3	0.1	5	0.9
Assayed cobble	5	0.1	11	2.1
Cobble hammerstone	2	0.1	2	0.3
Radial core/hammerstone	0	0	1	0.1
Multiplatform core/hammerstone	2	0.1	2	0.3
Bipolar artefact	31	0.8	?	–
Cobble manuport	0	0	1	0.1
Shatter fragment	0	0	33	6.5
Heat-fractured piece	12	0.3	0	0
Undiagnostic/identified artefact	0	0	49	9.6
Total	3,626	100	507	100

The data from Liang Bua are derived from sector IV, spits 23–85.

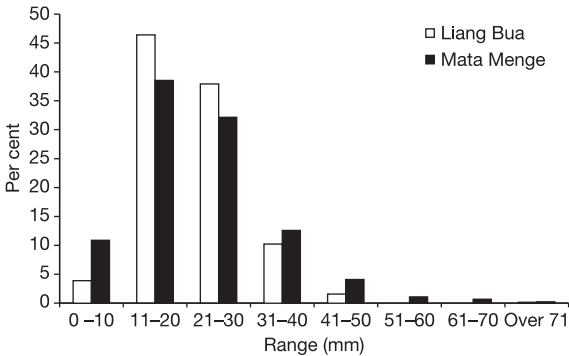


Figure 5 | Scar dimensions on Mata Menge and Liang Bua cores. The Liang Bua data come from spits 45–55, sector IV.

differences in sample size (Table 1)—a long-term technological theme involving the reduction of cores bifacially and radially, and the manufacture of a suite of technically and morphologically distinctive artefacts, is evident on Flores from at least 840 kyr ago right up to the disappearance of *H. floresiensis* 12 kyr ago^{8,9}. In contrast, the first skeletal evidence currently available for modern humans on the island, at Liang Bua around 10.5 kyr BP, is associated with various changes and additions to the stone artefact record, including an increased emphasis on the use of chert and the appearance of new stone artefact types (for example, edge-glossed flakes, grinding stones), as well as the first evidence for symbolic behaviour, such as personal ornaments (for example, beads), pigments and formal disposal of the dead⁹.

The stone artefact assemblages from Mata Menge and the Pleistocene levels of Liang Bua are remarkably similar. We still do not know the species identity of the Mata Menge knappers, as no associated hominin remains have been recovered so far, but the age of the site clearly precludes modern humans. At Liang Bua, however, the skeletal remains of at least nine individuals are represented in finds from the Pleistocene levels, and all diagnostic elements are of *H. floresiensis*⁹. The most parsimonious explanation for this is that the stone artefacts from Mata Menge and Liang Bua represent a continuous technology made by the same hominin lineage. Pronouncements that *H. floresiensis* lacked the brain size necessary to make stone artefacts¹⁷ are therefore based on preconceptions rather than actual evidence.

Received 1 August 2005; accepted 1 February 2006.

- O'Sullivan, P. B. *et al.* Archaeological implications of the geology and chronology of the Soa basin, Flores, Indonesia. *Geology* **29**, 607–610 (2001).
- Maringer, J. & Verhoeven, Th. Die steinartefakte aus der *Stegodon*-fossilschicht von Mengeruda auf Flores, Indonesien. *Anthropos* **65**, 229–247 (1970).
- Sondaar, P. Y. *et al.* Middle Pleistocene faunal turn-over and colonisation of Flores (Indonesia) by *Homo erectus*. *C.R. Acad. Sci.* **319**, 1255–1262 (1994).
- Morwood, M. J., O'Sullivan, P. B., Aziz, F. & Raza, A. Fission-track ages of stone tools and fossils on the east Indonesian island of Flores. *Nature* **392**, 173–176 (1998).
- Morwood, M. J. *et al.* Archaeological and palaeontological research in central Flores, east Indonesia: results of fieldwork 1997–98. *Antiquity* **73**, 273–286 (1999).
- Morwood, M. J., Aziz, F., van den Bergh, G. D., Sondaar, P. Y. & de Vos, J. Stone artefacts from the 1994 excavation at Mata Menge, west central Flores, Indonesia. *Aust. Archaeol.* **44**, 26–34 (1997).
- Lahr, M. M. & Foley, R. Human evolution writ small. *Nature* **431**, 1043–1044 (2004).
- Morwood, M. J. *et al.* Archaeology and age of a new hominin from Flores in eastern Indonesia. *Nature* **431**, 1087–1091 (2004).
- Morwood, M. J. *et al.* Further evidence for small-bodied hominins from the Late Pleistocene of Flores, Indonesia. *Nature* **437**, 1012–1017 (2005).
- Brown, P. *et al.* A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061 (2004).
- Falk, D. *et al.* The brain of LB1, *Homo floresiensis*. *Science* **308**, 242–245 (2005).
- van den Bergh, G. D. The Late Neogene elephantoid-bearing faunas of Indonesia and their palaeozoogeographic implications: a study of the terrestrial faunal succession of Sulawesi, Flores, and Java, including evidence for early hominid dispersal east of Wallace's Line. *Scripta Geol.* **117**, 1–419 (1997).
- Schick, K. D. *Stone Age Sites in the Making: Experiments in the Formation and Transformation of Archaeological Occurrences* (BAR International Series 319, Oxford, 1986).
- Shea, J. J. Artifact abrasion, fluvial processes, and "living floors" from the Early Paleolithic site of 'Ubeidiya (Jordan Valley, Israel). *Gearchaeology* **14**, 191–207 (1999).
- Maringer, J. & Verhoeven, Th. Notes on stone artefacts in the National Archaeology Institute of Indonesia at Djakarta, collected from the *Stegodon*-fossil bed at Boaleza in Flores. *Anthropos* **65**, 638–639 (1970).
- Moore, M. W. *The Design Space of Lithic Technology*. Thesis, Univ. New England (2005).
- Culotta, E. New 'hobbits' bolster species, but origins still a mystery. *Science* **310**, 208–209 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Excavations at Mata Menge, funded by an Australian Research Council Discovery grant to M.J.M., were undertaken from 26 August to 29 September 2004 and 16 August to 23 September 2005. The work was done under GRDC survey/research permission letter 57/45.04/BMG/2004 to the Governor of NTT Province, Kupang. We received support from the Ngadha Regency Administration and Head of the Bajawa Culture and Tourism Office, I. Yusuf. Other participants included P. M. D. Moi, I. Botha, E. Yan Patriani, S. Sudjarwadi and 20 Ngadha people led by K. Podhi. The Secretary of Desa Mengeruda, G. Leo, provided further support. We also thank Ngiliman and Dadang for preparing the topographic map of the Mata Menge area, and J. Tode Solo for assistance in the field. The Department of Archaeology and Natural History, Research School of Pacific and Asian Studies, Australian National University funded A.B. S. O'Connor and D. Boyd are thanked for their support.

Author Contributions A.B. conducted the analysis and interpretation of the Mata Menge assemblage. F.A. and G.D.v.d.B. planned and directed the excavations in association with M.J.M. G.D.v.d.B. described the stratigraphy and prepared Fig. 2. D.R.H. was responsible for aspects of the archaeological fieldwork and prepared Fig. 1. I.K. assisted with the palaeontological research. M.W.M. assisted with the Mata Menge analysis, conducted the Liang Bua analysis and prepared Figs 3–5. R.F. conducted the microwear analysis.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.B. (adam.brumm@anu.edu.au) or M.J.M. (mmorwood@pobox.une.edu.au).

Biodiversity and ecosystem stability in a decade-long grassland experiment

David Tilman¹, Peter B. Reich² & Johannes M. H. Knops³

Human-driven ecosystem simplification has highlighted questions about how the number of species in an ecosystem influences its functioning. Although biodiversity is now known to affect ecosystem productivity^{1–6}, its effects on stability are debated^{6–13}. Here we present a long-term experimental field test of the diversity–stability hypothesis. During a decade of data collection in an experiment that directly controlled the number of perennial prairie species⁴, growing-season climate varied considerably, causing year-to-year variation in abundances of plant species and in ecosystem productivity. We found that greater numbers of plant species led to greater temporal stability of ecosystem annual aboveground plant production. In particular, the decadal temporal stability of the ecosystem, whether measured with intervals of two, five or ten years, was significantly greater at higher plant diversity and tended to increase as plots matured. Ecosystem stability was also positively dependent on root mass, which is a measure of perenniating biomass. Temporal stability of the ecosystem increased with diversity, despite a lower temporal stability of individual species, because of both portfolio (statistical averaging) and overyielding effects. However, we found no evidence of a covariance effect. Our results indicate that the reliable, efficient and sustainable supply of some foods (for example, livestock fodder), biofuels and ecosystem services can be enhanced by the use of biodiversity.

The hypothesis that greater ecological diversity leads to greater stability⁷ has been a point of interest and debate for a half century^{7–14}. Field observations^{10,11,14,15} and laboratory experiments^{16–18} have generally shown that greater diversity is associated with greater ecosystem stability but lower species stability, much as predicted by models of multispecies competition^{8,19}. However, well-replicated field experiments that manipulate variables of interest have become indispensable in ecology. The only such field experiment to test for diversity–stability relations so far has been an eight-week study that had mixed results^{13,20}, leading a major review to conclude that the effects of diversity on stability remained unresolved⁶. We propose that diversity has consistent stabilizing effects on ecosystem processes once timescales are sufficient to incorporate the average net effects of diversity on both resistance to and recovery from perturbations.

Here we present the dependence of the temporal stability of ecosystems and species on plant diversity in a long-term grassland biodiversity experiment that established 168 plots containing 1–16 species⁴. Stability has several meanings in ecology^{8,21}. We focus on temporal stability, S , which measures the degree of constancy in a variable relative to its mean. S is defined as μ/σ , where μ is the mean value for a time period and σ is its temporal standard deviation over the same interval^{19,22}. Many factors can cause the abundances and primary productivity of plant species to vary, including precipitation^{10,23}, temperature²³, life histories, and interactions with other

organisms. During the ten years of data collection that followed two years of establishment for our 12-year experiment (Methods), precipitation during the growing season varied more than twofold and average daily high temperatures during the growing season ranged from 21.5 to 24.4 °C, creating growing seasons with widely different climatic conditions. Here we determine whether observed temporal variability in plant biomass was dependent on plant community diversity. We examine both ecosystem stability (temporal stability of aboveground plant biomass summed across all species in a plot, a measure of primary productivity) and species stability (temporal stability of the aboveground biomass of individual plant species).

We determined ecosystem stability with the use of all ten years (1996–2005) of data collected annually on aboveground biomass production within each plot. Ecosystem stability was determined both without detrending and after detrending data for each plot to remove variation attributable to a ten-year trend of generally increasing plot biomass in higher-diversity treatments (Methods). To ensure normality, all temporal stabilities were log transformed before analyses.

Detrended ecosystem stability was an increasing function of the number of planted species (Fig. 1a; $F_{1,161} = 44.9$, $P < 0.0001$). It was similarly dependent on the 1996–2005 plot-average Shannon diversity index, H' ($F_{1,161} = 13.7$, $P = 0.0003$), on plot-average effective species number, e^H ($F_{1,161} = 20.2$, $P < 0.0001$) and on realized species number (that is, the number of abundant species contributing to 90% of total biomass in 2005; Fig. 2; $F_{1,160} = 23.7$, $P < 0.0001$). Greater ecosystem stability at higher diversity meant there was lower proportional change in the annual production of biomass in plots with greater plant diversity. On average, the treatment plots with the highest diversity were about 70% more stable than monocultures (Fig. 1a). This greater ecosystem stability at higher diversity has been called the insurance value of biodiversity¹⁷.

We also measured ecosystem stability without detrending by dividing the ten years of data into non-overlapping intervals of shorter duration (two or five years) for which biomass had small or no temporal trends, and then calculating plot stability for each interval as $S = \mu/\sigma$. A repeated-measures multiple analysis of variance (MANOVA) that had the two five-year stabilities as dependent variables showed that greater numbers of planted species led to greater ecosystem stability ($F_{1,161} = 21.1$, $P < 0.0001$) and that stability was greater for the second five-year period (time effect: $F_{1,161} = 12.2$, $P = 0.0006$). A species-number $\times$ time interaction ($F_{1,161} = 6.29$, $P = 0.013$) indicated a stronger positive effect of greater species numbers on ecosystem stability in the second time period. Another repeated-measures MANOVA that had the five two-year measures of ecosystem stability as dependent variables showed that greater species numbers led to greater ecosystem stability

¹Department of Ecology, Evolution and Behavior, University of Minnesota, 1987 Upper Buford Circle, St Paul, Minnesota 55108, USA. ²Department of Forest Resources, University of Minnesota, 1530 North Cleveland Avenue, St Paul, Minnesota 55108, USA. ³School of Biological Sciences, University of Nebraska, 348 Manter Hall, Lincoln, Nebraska 68588-0118, USA.

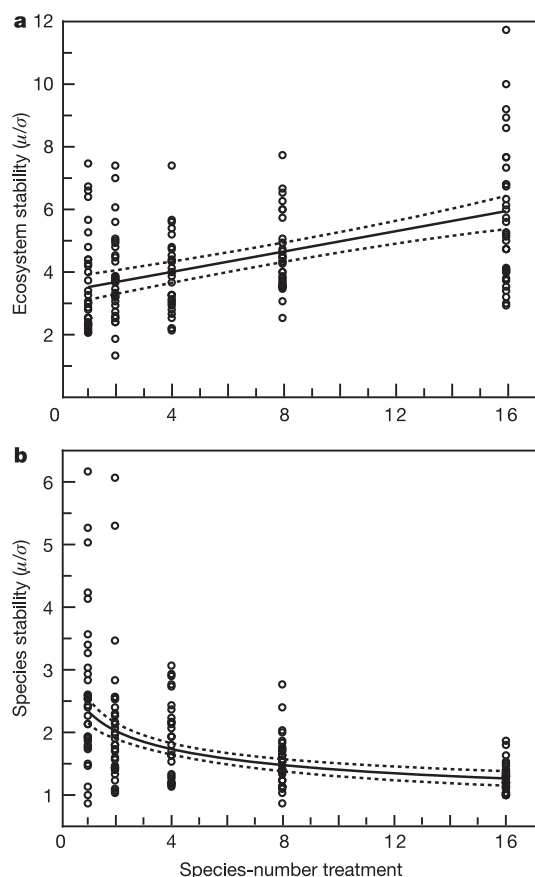


Figure 1 | Dependence of temporal stability of each plot on experimentally imposed species-number treatment. **a**, Ecosystem temporal stability for the decade from 1996 to 2005 was an increasing function of the number of planted species. Ecosystem stability is the ratio of mean plot total biomass to its temporal standard deviation, determined after detrending. The regression line and its 95% confidence interval are shown (untransformed data: $F_{1,159} = 43.7$, $P < 0.0001$). To reduce the difference in y axis scale between the two parts of this figure, a single data point (species number of 16, ecosystem stability of 15.76) is not shown but was included in all analyses. **b**, Plot-average species temporal stability, determined with species biomass data for 2001–2005, was a declining function of the number of planted species. The regression curve and 95% confidence intervals are based on a fit of $\log(\text{species stability})$ on $\log(\text{species number})$, with $F_{1,159} = 72.3$, $P < 0.0001$.

($F_{1,155} = 16.5$, $P < 0.0001$), stability had a weak tendency to increase through time ($F_{4,152} = 2.24$, $P = 0.067$) and there was no species-number $\times$ time interaction ($F_{4,152} = 0.60$, $P = 0.66$). Similar repeated-measures MANOVAs, of both two-year and five-year stabilities, that used realized species number as the independent variable yielded similar results. The greater ecosystem stability of higher-diversity plots resulted from their having lower temporal standard deviations, for a given mean plot biomass, than plots with lower diversity (Fig. 3). In total, on average across the decade of measurement, ecosystem stability was significantly positively dependent on plant diversity, and this result was robust with respect to data detrending and the intervals over which stability was determined.

In contrast to ecosystem stability, stabilities of individual species (log transformed), determined with our five-year record of abundances of each species planted in each plot, were a declining function of the number of planted species ($F_{1,988} = 134.3$, $P < 0.0001$) and, similarly, of effective species number, e^H ($F_{1,988} = 83.6$, $P < 0.0001$). We also calculated the average, for each plot, of the species stabilities of all species planted in the plot, and found that the plot-average

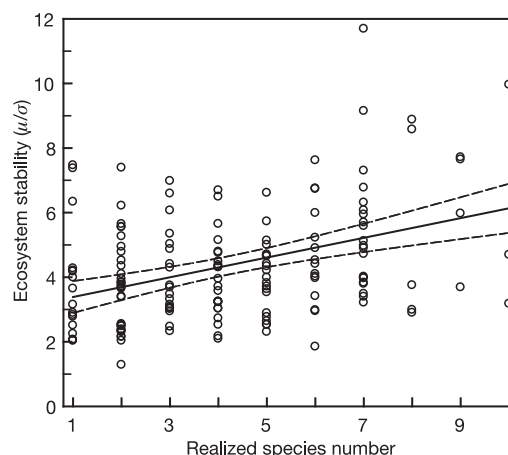


Figure 2 | Dependence of ecosystem temporal stability from 1996 to 2005 on realized species number. All species, whether planted or weedy, were ranked by proportional abundance in the sorted 0.6 m² clipped strip of 2005. Proportional abundances were summed, in order from the most abundant species, to determine the realized species number, which is the number of more abundant species comprising 90% of the total aboveground biomass of a plot. Ecosystem stability was also significantly dependent on realized species number determined with cutoffs of 75% ($P < 0.0001$) and 99% ($P = 0.002$). As in Fig. 1a, one data point (realized species number of 7.2, ecosystem stability of 15.76) is not shown but was included in all analyses.

species stability (log transformed) was a declining function of planted species number ($F_{1,159} = 63.5$, $P < 0.0001$; Fig. 1b) and of e^H ($F_{1,159} = 57.1$, $P < 0.0001$). Species stabilities were not detrended, but results were similar if detrended.

We used stepwise regression to evaluate the influence of root mass, functional group composition (presence or absence of C₃ grasses, C₄ grasses, legumes or non-legume forbs), weedy species biomass, initial soil fertility (initial total soil nitrogen) and species number on ecosystem stability. In both forward addition and backward elimination analyses, the same three variables were retained, with ten-year detrended ecosystem stability remaining positively dependent on species number ($F_{1,159} = 16.2$, $P < 0.0001$) and also being positively dependent on root mass ($F_{1,159} = 23.0$, $P < 0.0001$) but negatively dependent on the presence of legumes ($F_{1,159} = 4.42$, $P = 0.037$). The positive effect of root mass probably occurred because roots are the perennating structure of these herbaceous perennial species, and higher root mass should provide a larger store of nutrients and energy to buffer growth in response to environmental variation. Weedy biomass had no significant ($P > 0.05$) effects on stability and was neither added nor retained in the forward or backward stepwise regressions, respectively. Similarly, in repeated measures analyses using two-year or five-year ecosystem stabilities, weed biomass had no significant effects ($P > 0.1$) but ecosystem stability remained an increasing function of numbers of species planted ($P < 0.001$). This indicates that any disturbance that might have been associated with differences between treatments in weeding intensity did not influence results. Diversity did affect invading weedy species. After cessation of weeding in subplots, total numbers of plant species and total biomass increased more at lower diversity than at higher diversity²⁴.

The strength and consistency of the long-term stabilizing effects of diversity on ecosystem productivity that we observed contrast with mixed effects observed when a short-term drought was imposed on a biodiversity experiment¹³. In that study, the proportion of aboveground plant biomass lost after an 8-week drought was independent of diversity¹³, indicating, by a metric analogous to ours, no effect of diversity on short-term proportional resistance stability. Because more diverse plots had greater biomass, the absolute biomass loss was greater at greater diversity, which was interpreted as showing lower absolute resistance stability at higher diversity¹³. However, during the

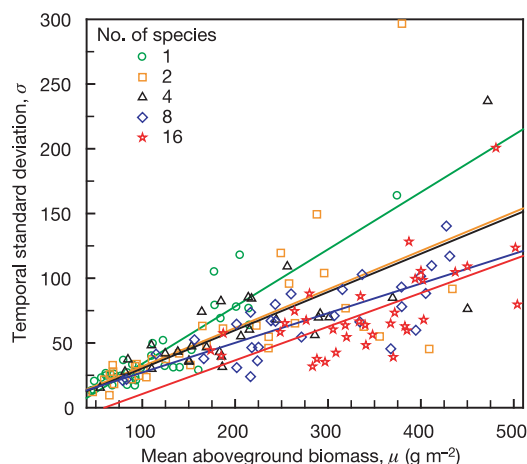


Figure 3 | Effects of plant diversity on the relationship between mean biomass and its temporal standard deviation. Each point shows the mean biomass of a plot and its temporal standard deviation based on the five annual samples collected from 2001 to 2005. Data for each species-number treatment were fitted by a separate regression line. Data were not detrended. The results show that high-diversity treatments have lower temporal standard deviations (lower risks) for a given mean biomass (return).

four years before imposition of drought, more diverse plots had greater temporal stability²⁵. In combination, these two studies indicate that there might be short-term variability in effects of diversity on ecosystem stability but that, in the long term, higher plant diversity causes greater ecosystem stability.

Our results support the predictions of competition theory that greater diversity leads to greater ecosystem stability and lower species stability^{8,19,22}. This theory predicts that greater ecosystem stability at higher diversity can result from increasingly negative covariance in the abundances of competing species at higher diversity^{19,22} (covariance effect), from the manner in which temporal variance in species abundances scales with abundance^{19,22,26,27} (statistical averaging or portfolio effect), and/or from the manner in which species abundances scale with diversity^{19,22} (overyielding effect; greater ecosystem total biomass at higher diversity).

The covariance effect requires that total covariance (temporal covariance in abundances for each pair of species, summed across all possible pairs of species) decline as diversity increases. However, regression showed no dependence of total covariance for 2001–05 on species number ($F_{1,161} = 0.83$, $P = 0.36$). Our results thus do not support the covariance effect, a conclusion also reached in an earlier analysis of non-experimental data²². The portfolio effect^{22,26} requires temporal variance, s^2 , in the abundance of a species to scale with its biomass, m , as $s^2 = c_1 m^z$, with $1 < z < 2$, which it did, with $z = 1.60$. This supports the role of the portfolio effect in stabilizing higher-diversity plots. In multiple regression, summed variances were a declining function of species number ($F_{1,160} = 15.7$, $P < 0.0001$) and an increasing function of total biomass ($F_{1,160} = 127$, $P < 0.0001$), which also supports the portfolio effect. The overyielding effect requires²² that total plot biomass, B , increase with diversity, which it did each year of the experiment. For 2001–05, plots containing 16 species had, on average, 180% more biomass than monocultures. In total, these analyses strongly indicate that greater ecosystem stability at higher diversity resulted from portfolio and overyielding effects.

Rapidly increasing human population and consumption, and concomitant demand for food and energy, are making society increasingly dependent on services provided by remaining natural and managed ecosystems^{28,29}. Biodiversity experiments have shown that greater numbers of plant species lead to a greater production of biomass^{1–6}. Our results show that the long-term stability of an

ecosystem service—the annual production of biomass and thus of potential biofuels and livestock fodder^{1–6}—also depends on biodiversity. Biodiversity can therefore be an important element for the reliable and sustainable provisioning of ecosystem services.

METHODS

Experimental design. In a 7-ha field at Cedar Creek Natural History Area, Minnesota, USA, we controlled the number of plant species in 168 plots, each 9 m × 9 m. Plots were randomly assigned to be seeded with 1, 2, 4, 8 or 16 perennial grassland species, with 39, 35, 29, 30 and 35 replicates, respectively, of the diversity levels. The composition of each plot was randomly chosen from a set of 18 perennials (four C_4 grasses, four C_3 grasses, four legumes, four non-legume forbs and two woody species). All plots received 10 g m⁻² of seed in May 1994 and 5 g m⁻² in May 1995, with seed mass divided equally between species. Treatments were maintained by weeding three or four times each year. Weeds were removed while still small, with care being taken to minimize any disturbance. Plots were burned annually in spring before growth began. Five woody monocultures are not included in analyses because burning effectively eliminated woody species from multispecies plots. Plots were annually sampled in mid-August for aboveground living plant biomass by clipping, drying and weighing four parallel and evenly spaced 0.1 m × 3.0 m vegetation strips per plot from 1996 to 1999 and four 0.1 m × 6.0 m strips per plot from 2000 to 2005. Different locations were clipped each year. Biomass from one strip per plot was sorted to species from 2001 to 2005. The Shannon diversity index, H' , used abundances of each species, planted or weedy, in each plot by means of estimates of percentage cover for 1996–2000 (four 0.5 m² subplots per plot) and sorted biomass for 2001–2005. See ref. 4 for further details.

Sampling effort. To eliminate potential bias from different sampling efforts for the first four in comparison with the last six years, for each of the last six years two clipped strips per plot were randomly chosen for an analysis of ecosystem stability. The full data gave similarly significant and positive effects of diversity on all three measures of ecosystem stability.

Detrending and other analyses. Detrending was done, for each plot, by means of linear regression of annually measured plot biomass on the logarithm of year and used all ten years of plot data. The logarithm of year provided a generally better fit than year; both gave similar results. The standard deviation, σ_d , of residuals for each regression measures detrended variation. The detrended temporal stability, S_d , of a plot was $S_d = \mu/\sigma_d$. Each plot had a single detrended stability value for the ten-year period. In contrast, when data were divided into shorter intervals that did not require detrending, there were multiple values of S , calculated as $S = \mu/\sigma$, per plot. We divided the data either into two subsets, each five years in duration (1996–2000 and 2001–05) or into five subsets, each two years in duration (1996–97, 1998–99, 2000–01, 2002–03 and 2004–05). These temporal sequences of S values for each plot were analysed with the use of repeated-measures MANOVA.

Received 20 December 2005; accepted 23 March 2006.

- Naeem, S., Häkenson, K., Lawton, J. H., Crawley, M. J. & Thompson, L. J. Biodiversity and plant productivity in a model assemblage of plant species. *Oikos* **76**, 259–264 (1996).
- Tilman, D., Wedin, D. & Knops, J. Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* **379**, 718–720 (1996).
- Hector, A. *et al.* Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127 (1999).
- Tilman, D. *et al.* Diversity and productivity in a long-term grassland experiment. *Science* **294**, 843–845 (2001).
- Hector, A., Bazeley-White, E., Loreau, M., Otway, S. & Schmid, B. Overyielding in grassland communities: Testing the sampling effect hypothesis with replicated biodiversity experiments. *Ecol. Lett.* **5**, 502–511 (2002).
- Hooper, D. U. *et al.* Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol. Monogr.* **75**, 3–35 (2005).
- Elton, C. S. *The Ecology of Invasions by Animals and Plants* (Methuen, London, 1958).
- May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, Princeton, 1973).
- McNaughton, S. J. Stability and diversity of ecological communities. *Nature* **274**, 251–253 (1978).
- Tilman, D. & Downing, J. A. Biodiversity and stability in grasslands. *Nature* **367**, 363–365 (1994).
- Tilman, D. Biodiversity: Population versus ecosystem stability. *Ecology* **77**, 350–363 (1996).
- Huston, M. A. Hidden treatments in ecological experiments: Re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).
- Pfisterer, A. B. & Schmid, B. Diversity-dependent production can decrease the stability of ecosystem functioning. *Nature* **416**, 84–86 (2002).

14. McNaughton, S. J. in *Biodiversity and Ecosystem Function* (eds Schulze, E.-D. & Mooney, H. A.) 361–383 (Springer, Berlin, 1993).
15. Frank, D. A. & McNaughton, S. J. Stability increases with diversity in plant communities: Empirical evidence from the 1988 Yellowstone drought. *Oikos* **62**, 360–362 (1991).
16. McGrady-Steed, J., Harris, P. M. & Morin, P. J. Biodiversity regulates ecosystem predictability. *Nature* **390**, 162–165 (1997).
17. Naeem, S. & Li, S. Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509 (1997).
18. Petchey, O. L., McPhearson, P. T., Casey, T. M. & Morin, P. J. Environmental warming alters food-web structure and ecosystem function. *Nature* **402**, 69–72 (1999).
19. Lehman, C. L. & Tilman, D. Biodiversity, stability, and productivity in competitive communities. *Am. Nat.* **156**, 534–552 (2000).
20. Naeem, S. Biodiversity equals instability? *Nature* **416**, 23–24 (2002).
21. Pimm, S. L. The complexity and stability of ecosystems. *Nature* **307**, 321–326 (1984).
22. Tilman, D. The ecological consequences of changes in biodiversity: A search for general principles. *Ecology* **80**, 1455–1474 (1999).
23. Tilman, D. & El Haddi, A. Drought and biodiversity in grasslands. *Oecologia* **89**, 257–264 (1992).
24. Fargione, J. E. & Tilman, D. Diversity decreases invasion via both sampling and complementarity effects. *Ecol. Lett.* **8**, 604–611 (2005).
25. Pfisterer, A. B., Joshi, J., Schmid, B. & Fischer, M. Rapid decay of diversity–productivity relationships after invasion of experimental plant communities. *Basic Appl. Ecol.* **5**, 5–14 (2004).
26. Doak, D. F. et al. The statistical inevitability of stability–diversity relationships in community ecology. *Am. Nat.* **151**, 264–276 (1998).
27. Tilman, D., Lehman, C. L. & Bristow, C. E. Diversity–stability relationships: Statistical inevitability or ecological consequence? *Am. Nat.* **151**, 277–282 (1998).
28. Daily, G. C. *Nature's Services: Societal Dependence on Natural Ecosystems* (Island Press, Washington DC, 1997).
29. Kremen, C. Managing ecosystem services: What do we need to know about their ecology? *Ecol. Lett.* **8**, 468–479 (2005).

Acknowledgements We thank J. Fargione for insightful comments, and T. Mielke, N. Larson, S. Bauer and our summer interns for assistance. The National Science Foundation, the Bush Foundation and the Andrew Mellon Foundation provided financial support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.T. (tilman@umn.edu).

Frequency-dependent survival in natural guppy populations

Robert Olendorf¹, F. Helen Rodd³, David Punzalan³, Anne E. Houde⁴, Carla Hurt⁵, David N. Reznick⁶ & Kimberly A. Hughes^{1,2}

The maintenance of genetic variation in traits under natural selection is a long-standing paradox in evolutionary biology^{1–3}. Of the processes capable of maintaining variation, negative frequency-dependent selection (where rare types are favoured by selection) is the most powerful, at least in theory¹; however, few experimental studies have confirmed that this process operates in nature. One of the most extreme, unexplained genetic polymorphisms is seen in the colour patterns of male guppies (*Poecilia reticulata*)^{4,5}. Here we manipulated the frequencies of males with different colour patterns in three natural populations to estimate survival rates, and found that rare phenotypes had a highly significant survival advantage compared to common phenotypes. Evidence from humans^{6,7} and other species^{8,9} implicates frequency-dependent survival in the maintenance of molecular, morphological and health-related polymorphisms. As a controlled manipulation in nature, this study provides unequivocal support for frequency-dependent survival—an evolutionary process capable of maintaining extreme polymorphism.

Colour-pattern polymorphism in guppies is limited to males and consists of irregular spots of several different structural (blue, green and purple) and pigment-based (yellow, orange, red and black) colours that occur on the body, caudal fin and dorsal fin (Fig. 1). The position, number, size and hue of the spots are highly heritable^{5,10}, although the colour saturation (chroma) of orange spots can be influenced by diet^{11,12}. Male colouration is highly polymorphic despite being subject to sexual and ecological selection. Female mating preferences usually favour males with the greatest area of orange, although the strength of that preference varies among populations^{10,12,13}. Predators also exert selection on colour patterns; they preferentially prey upon males with brighter or more conspicuous colours^{14,15}. Despite the apparently strong and directional selection within populations, colour patterns are so variable that any two males are easily distinguishable based on colour pattern alone, unless they are closely related¹⁰.

Several mechanisms have been proposed to explain the maintenance of this extreme polymorphism^{10,14}. Mate-choice experiments indicate that females preferentially mate with males bearing rare or novel colour patterns^{16,17}. A trade-off (antagonistic pleiotropy) between male sexual attractiveness and offspring viability has also been reported¹⁸. Both processes could contribute to the maintenance of genetic variation in nature. However, experiments demonstrating these processes were conducted in laboratory environments, and it is not clear whether either process occurs in nature. Another process capable of maintaining polymorphism is a rare-morph survival advantage. This process has been implicated in the maintenance of colour polymorphism in some invertebrates^{19,20} and vertebrates²¹, but it has not been tested in the highly polymorphic guppy system.

We tested the hypothesis that male survival is causally related to colour-pattern rareness in three natural guppy populations in Trinidad. We used an established mark–release–recapture protocol^{22–24} to estimate the survival of wild guppies in native streams, where we manipulated the frequency of male colour patterns. Under frequency-dependent selection, genotypes have equal fitness when they are at equilibrium frequencies; therefore, we manipulated the frequencies of different morphs to move the population away from equilibrium. We conducted 34 separate manipulations across 19 replicate pools in three streams over four years. If frequency-dependent selection occurs, we expected a survival advantage for rare phenotypes relative to common phenotypes.

We conducted these experiments in two unconnected tributaries of the Quare River (Quare 1 and Quare 7), and the main branch of the Mausica River. Both Quare tributaries are small pool-and-riffle streams where the dominant predator is a killifish (*Rivulus hartii*). The Mausica River is a larger stream with pool–riffle topology, having both *R. hartii* and the pike cichlid (*Crenicichla alta*), which is thought to be the dominant predator where it occurs. *C. alta* prey predominantly on adult female and male guppies, and *R. hartii* consume juvenile and adult male guppies, but not large adult females^{23,25}. These piscivorous fish are the only major predators of guppies in the sites we used²³. Guppies show little intraspecific aggression, and intraspecific interactions are not a direct source of mortality¹⁰. Consequently, predators are the most likely sources of short-term mortality for guppies. In each stream, we used four to seven pools of similar size and structure in each of two different years. All replicate pools within a stream were similar with respect to substrate and water clarity. The presence of adult *R. hartii* and/or *C. alta* was confirmed for each pool used in the experiment.

After collecting all adult males and females from each experimental pool at a site, the males from all pools within a site were combined and then sorted based on the classification of alternate tail colour patterns, which were nearly equal in abundance within a site. We used tail colour patterns because they are distinctive, likely to be conspicuous to predators, and are assumed to have an important role in courtship. For convenience, alternate morphs were designated as ‘Coloured’ and ‘Uncoloured’, or ‘Flag’ and ‘Blob’, depending on the site and year (Fig. 1). In half of the replicates within a site, one morph was made rare and the other common, typically in a ratio of 3:1. In the remainder of the pools, rare and common morph frequencies were reversed (Supplementary Table S1). Within a morph category, males were randomly assigned to experimental pools. We released adult females into the same pool from which they had been collected, and maintained natural densities and sex ratios in each pool. All adults were given a pool-specific mark^{23,24} so that any migrants could be identified.

¹School of Integrative Biology and ²Institute for Genome Biology, University of Illinois, Urbana, Illinois 61801, USA. ³Department of Zoology, University of Toronto, Toronto, Ontario M5S 3G5, Canada. ⁴Department of Biology, Lake Forest College, Lake Forest, Illinois 60045, USA. ⁵Smithsonian Tropical Research Institute, Naos Marine Laboratory, Roosevelt Avenue (APO AA 34002), Panama City, Panama. ⁶Department of Biology, University of California, Riverside, California 92521, USA.

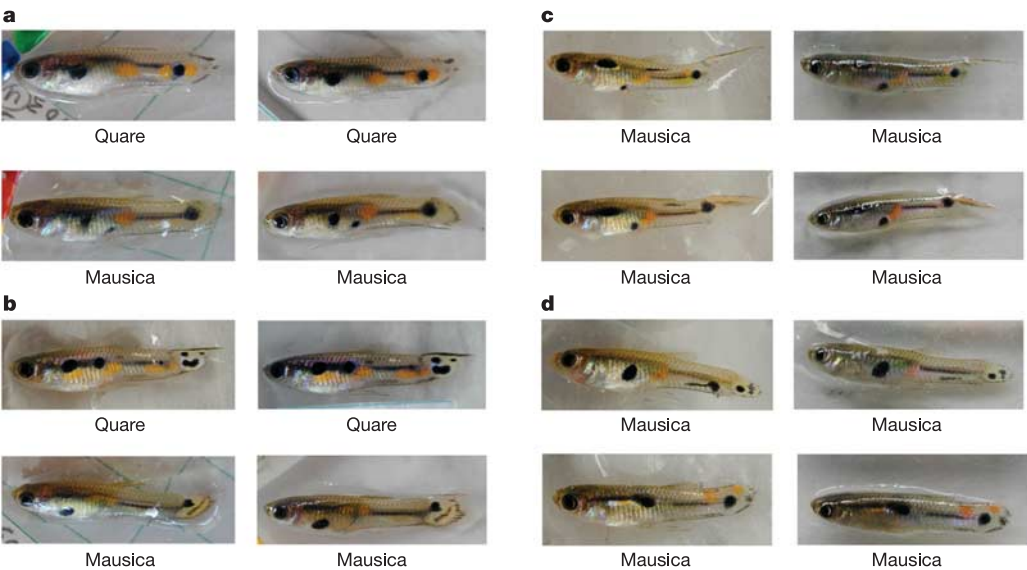


Figure 1 | Males before release and after recapture. In each panel, photographs taken before release are on the left; photographs of the same animal after recapture are on the right. Variations in hue reflect ambient light variation in the field. Tattoo marks are visible as thin horizontal lines on the caudal peduncles. **a, b**, Examples of ‘Uncoloured’ (**a**) and ‘Coloured’

(**b**) morphs; **c, d**, examples of ‘Flag’ (**c**) and ‘Blob’ (**d**) morphs (see Supplementary Methods). Panel **c** illustrates that even males with very similar colour patterns can be distinguished when matching before and after photos.

The release phase of the experiment lasted 15 (Mausica) or 17 (Quare tributaries) days. At the end of this phase, we again exhaustively searched until we had captured all adult-size guppies from each pool, and from non-experimental pools up- and downstream to identify any migrants. We classified recaptured males with respect to colour pattern and pool-specific mark, and compared photographs of released and recaptured males to determine the recapture status of individual males (Fig. 1).

Males with the rare phenotype had significantly higher recapture rates overall than did males with the common phenotype (Table 1 and Fig. 2), suggesting that rare types had a large survival advantage. Rare types also had higher recapture rates within each site; site-specific contrasts were significant in Quare 1 ($\chi^2 = 461$; degrees of freedom (d.f.) = 1; $P < 0.0001$) and Quare 7 ($\chi^2 = 347$; d.f. = 1; $P < 0.0001$), with a similar trend in Mausica ($\chi^2 = 3.30$; d.f. = 1; $P = 0.07$) (Fig. 2). Notably, we recaptured every rare male in every pool in Quare 7 (1996) and in Quare 1 (2004), whereas only 61% and 67% of common males were recaptured, respectively (Table 2). Repeated-measures analysis of variance (ANOVA) of recapture rates, treating rare and common morphs within a pool and year as repeated measures, yielded similar results (see Supplementary Notes). Rare types had higher recapture rates within years at each site, except for one site–year combination (Quare 7, 1999). Therefore, this site accounts for the significant site $\times$ rarity $\times$ year interaction effect in Table 1; when the data from this site are removed, the effect of rarity is still significant ($\chi^2 = 11.23$; d.f. = 1; $P = 0.0008$), but the interaction term is no longer significant ($\chi^2 = 5.8$; d.f. = 2; $P = 0.06$).

Table 1 | Model effects and likelihood-ratio tests

Effect	Degrees of freedom	χ^2	P
Site	2	3.61	0.16
Frequency	1	14.73	0.0001
Morph	1	0.70	0.40
Site $\times$ frequency	2	2.54	0.28
Site $\times$ year	2	10.35	0.006
Frequency $\times$ morph	2	0.01	0.93
Site $\times$ frequency $\times$ year	3	17.69	0.0005
Site $\times$ morph $\times$ year	5	6.14	0.29

Morphs did not differ in recapture rates overall or within sites (Table 1). Thus, there is no evidence of differential survival between the phenotypes we used. Total recapture rate was 73%, in accordance with estimates from other studies^{22,23}. There were no morph $\times$ frequency interactions, indicating that the particular phenotype that was rare or common had no effect on recapture probability.

We reanalysed the data to determine whether any bias was introduced by the few males who migrated or by the unmarked males that were captured at the end of the experiment in some pools. A few marked males at Quare 7 migrated to other pools before recapture; these males were counted as survivors in the above analyses. However, if the males migrated early in the release phase, they would not have experienced the same morph–frequency environment as other males in their experimental pool. We therefore removed these males from the data set. This change had little effect on the results (Supplementary Tables S2 and S3). We also repeated the analysis after excluding four samples in which unmarked males could have altered morph frequencies substantially (Mausica 2003 pools A and D, and 2004 pool D; Quare 7 1996 pool 12). Again, results were essentially unchanged (Supplementary Table S2). We

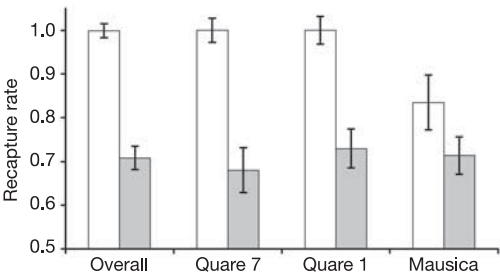


Figure 2 | Relative recapture rates for rare and common phenotypes. Least-square mean recapture rates and standard errors for rare (open bars) and common (grey bars) phenotypes were calculated from the full model, and are therefore corrected for other terms in the model. The vertical axis should be interpreted as relative, not absolute, survival. Over all sites, the difference in recapture rates between rare and common morphs is significant. Within sites, the difference is significant for Quare 7 and Quare 1 (see the main text).

Table 2 | Recapture rates by population and year*

Site	Year	Rare	Common
Mausica	2003	0.88 (0.08)	0.72 (0.04)
Mausica	2004	0.79 (0.12)	0.66 (0.08)
Quare 1	2003	0.79 (0.07)	0.75 (0.04)
Quare 1	2004	1.00 (0.00)	0.67 (0.10)
Quare 7	1996	1.00 (0.00)	0.61 (0.10)
Quare 7	1999	0.60 (0.11)	0.74 (0.08)

*Values for rare and common morphs are means (and standard errors).

evaluated the potential for bias due to the random assignment of some males to the same pool from which they were initially captured, and found that accounting for a 'home pool' effect did not substantially change the results (Supplementary Notes).

In summary, this experiment demonstrates that frequency-dependent survival occurs in natural guppy populations, and supports the hypothesis that frequency-dependence contributes to the extreme polymorphism of male colour patterns. Although we have not measured the evolutionary response to selection in this study, the high heritability of guppy colour patterns^{4,5,10} suggests that short-term response to selection would be strong. Selection on other components of fitness could oppose survival selection on colour patterns, but the available evidence indicates that differential reproduction also favours rare colour patterns^{16,17,26} because females have a mating preference for rare or novel males¹⁷. An intriguing possibility is that frequency-dependent survival could lead to direct selection for a rare-male mating preference if the reduction in risk extends to females involved in courtship.

Given that the known guppy predators hunt visually, a possible mechanistic explanation for a survival advantage of rare morphs is that predators form a 'search image' for common morphs, resulting in an increased ability to detect familiar prey but a decreased probability of detecting alternative prey. Search image formation is thought to result from limited attention in predators; it can generate frequency-dependent predation and maintain polymorphism, as shown in experiments using artificial prey^{27,28}. If selective predation were responsible for the frequency-dependent survival we observed, our results suggest that predators were attuned to small elements of colour patterns, as we manipulated the frequency of only a small part of the overall pattern. An alternative to the search image hypothesis is that male guppies altered their behaviour in response to morph frequency, and that differential survival was related to these behavioural changes. Behavioural tests on both predators and prey will be needed to discriminate between these hypotheses, and to determine whether predators distinguish between small elements of the overall colour pattern.

METHODS

Mark–release–recapture experiments. We conducted mark–release–recapture experiments at three sites (the exact locations are given in Supplementary Notes) in different years, using a protocol that has been used successfully to estimate survival in natural guppy populations in previous investigations^{22–24}. At each site, we used adjacent or nearly adjacent pools occurring within a single section of stream (<750 m). Before each experiment, we exhaustively collected all adult and subadult guppies from each pool. Guppies were clearly visible because all pools had shallow, clear water and no aquatic vegetation. We returned to the same pools on 2–3 consecutive days until no fish were collected during 45 min of searching at each pool. Captured guppies were separated by sex and kept in aquaria for 2–8 days in our field laboratory. We marked anaesthetized fish with a small subcutaneous elastomer tattoo (Northwest Marine Technology) unique to each experimental pool. We used only white or black tattoos to avoid colours implicated in female choice¹⁴; the colour of the marks does not affect survival²³.

In a given year, males collected from different pools within a site were combined and sorted on the basis of colour pattern. In all but one replicate, males were sorted into 'Coloured' and 'Uncoloured' phenotypes (Fig. 1). In Mausica 2004, males with >50% of the caudal fin coloured were rare, so a different classification was used: 'Flag' or 'Blob' (Fig. 1; Supplementary Methods). We did not use males not clearly falling into either category, or

young males without fully developed colour patterns. Within a colour category, males were randomly assigned to release pools at the frequencies shown in Supplementary Table 1. Slight variation among pools occurred when there were insufficient numbers of one type. Each male was digitally photographed before release but after marking. We released a total of 459 marked males and recaptured 312 marked males. Fish were not mixed across sites, and were not used in multiple years within a site.

For each site in each year, equal numbers of adult females were reintroduced into all pools (8–25 females per pool), chosen to retain pre-experimental population density and sex ratio. Releases were made a minimum of 24 h (usually 2–3 days) after marking, and after removing any adult guppies that were found in experimental pools. Before release, we built nylon screen barriers at the up- and downstream ends of each pool to reduce migration of guppies into or out of the pool during the release phase of the experiment.

On recapture, we anaesthetized males and photographed them again. Three independent individuals matched 'before' and 'after' photographs (see Fig. 1). Any male that had a 'before' photo but not an 'after' photo was assumed to have died during the release phase. It is possible that a small number of the males managed to bypass our barriers and migrated beyond the pools where we checked for migrants. However, there was no significant difference between rare and common morphs in the few marked migrant males that we did capture (all in Quare 7): 3/29 rare versus 7/86 common (likelihood ratio $\chi^2 = 0.11$; $P = 0.74$). The slight tendency for higher migration rates in rare males would have led us to underestimate rare male survival. In order to affect our conclusions, there would have to be a rare-male bias in long-distance migration, but not in short-distance migration—a pattern that seems unlikely and for which there is no evidence.

Statistics. The proportions of recaptured males in each category in each pool were analysed using SAS Proc Genmod with logit link, binomial distribution function, and type 3 likelihood ratio tests²⁹. This procedure provides a generalized linear modelling framework for the analysis of categorical data. We fitted the following model of categorical effects: $\text{logit}(\pi_{ijkl}) = \mu + \alpha_i + \beta_j + \delta_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{il} + (\beta\delta)_{jk} + (\alpha\beta\gamma)_{ijl} + (\alpha\delta\gamma)_{ikl}$, where π_{ijkl} is the proportion of males recaptured at site i , in rarity category j and morph category k in year l . Parentheses denote interaction terms. This model passed goodness-of-fit tests with no evidence of overdispersion (Pearson $\chi^2 = 59.0$; d.f. = 49; $P > 0.15$), and additional terms did not improve goodness of fit. We used CONTRAST statements to test the significance of recapture rates between frequency categories within sites. Means and standard errors of recapture rates for sites and years (Table 2) were calculated by averaging over the pool-specific proportions from that site and year. Model-corrected means (Fig. 2) were calculated using inverse logit transformations, $(\bar{\pi} = e^{\bar{\eta}} / (1 + e^{\bar{\eta}}))$. Standard errors (s.e.) were determined using the delta method, $\text{s.e.}(\bar{\pi}) = \bar{\pi}(1 - \bar{\pi}) \times \text{s.e.}(\bar{\eta})$; because the delta method produces standard error estimates of zero when $\bar{\eta} = 1$, we replaced values of $\bar{\eta} > 0.95$ with $\bar{\eta} = 0.95$ to calculate approximate standard errors. Parametric linear models were fitted to arc-sine square-root-transformed proportions.

Received 25 October 2005; accepted 9 February 2006.

- Lewontin, R. C. *The Genetic Basis of Evolutionary Change* (Columbia Univ. Press, New York, 1974).
- Charlesworth, B. & Hughes, K. A. in *Evolutionary Genetics from Molecules to Morphology* (eds Singh, R. S. & Krimbas, C. B.) 369–391 (Cambridge Univ. Press, Cambridge, 2000).
- Turelli, M. & Barton, N. H. Polygenic variation maintained by balancing selection: pleiotropy, sex-dependent allelic effects and GxE interactions. *Genetics* **166**, 1053–1079 (2004).
- Hughes, K. A., Rodd, F. H. & Reznick, D. N. Genetic and environmental effects on secondary sex traits in guppies (*Poecilia reticulata*). *J. Evol. Biol.* **18**, 35–45 (2005).
- Brooks, R. & Endler, J. A. Direct and indirect sexual selection and quantitative genetics of male traits in guppies (*Poecilia reticulata*). *Evolution* **55**, 1002–1015 (2001).
- Trachtenberg, E. et al. Advantage of rare HLA supertype in HIV disease progression. *Nature Med.* **9**, 928–935 (2003).
- Borghans, J. A. M., Beltman, J. B. & De Boer, R. J. MHC polymorphism under host–pathogen coevolution. *Immunogenetics* **55**, 732–739 (2004).
- Reid, D. G. Natural selection for apostasy and crypsis acting on the shell colour polymorphism of a mangrove snail, *Littoraria filosa* (Sowerby) (Gastropoda: Littoriniidae). *Biol. J. Linn. Soc.* **30**, 1–24 (1987).
- Sinervo, B., Bleay, C. & Adamopoulou, C. Social causes of correlational selection and the resolution of a heritable throat color polymorphism in a lizard. *Evolution* **55**, 2040–2052 (2001).
- Houde, A. E. *Sex, Color and Mate Choice in Guppies* (Princeton Univ. Press Monographs in Behavioral Ecology, Princeton, 1997).
- Kodric-Brown, A. Dietary carotenoids and male mating success in the guppy: an environmental component to female choice. *Behav. Ecol. Sociobiol.* **25**, 393–401 (1989).

12. Grether, G. F. Carotenoid limitation and mate preference evolution: a test of the indicator hypothesis in guppies (*Poecilia reticulata*). *Evolution* **54**, 1712–1724 (2000).
13. Endler, J. A. & Houde, A. E. Geographic variation in female preferences for male traits in *Poecilia reticulata*. *Evolution* **49**, 456–468 (1995).
14. Endler, J. A. Natural selection on color pattern in *Poecilia reticulata*. *Evolution* **34**, 76–91 (1980).
15. Godin, J.-G. J. & McDonough, H. E. Predator preference for brightly colored males in the guppy: a viability cost for a sexually selected trait. *Behav. Ecol.* **14**, 194–200 (2003).
16. Farr, J. A. Male rarity or novelty, female choice behavior and sexual selection in the guppy *Poecilia reticulata* Peters (Pisces: Poeciliidae). *Evolution* **31**, 162–168 (1977).
17. Hughes, K. A., Du, L., Rodd, F. H. & Reznick, D. N. Familiarity leads to female mate preference for novel males in the guppy, *Poecilia reticulata*. *Anim. Behav.* **58**, 907–916 (1999).
18. Brooks, R. Negative genetic correlation between male sexual attractiveness and survival. *Nature* **406**, 67–70 (2000).
19. Jones, J. S., Leith, B. H. & Rawlings, P. Polymorphism in *Cepaea*: a problem with too many solutions? *Annu. Rev. Ecol. Syst.* **8**, 109–143 (1977).
20. Oxford, G. S. & Gillespie, R. G. Evolution and ecology of spider coloration. *Annu. Rev. Entomol.* **43**, 619–643 (1998).
21. Sinervo, B., Svensson, E. & Comendant, T. Density cycles and an offspring quantity and quality game driven by natural selection. *Nature* **406**, 985–988 (2000).
22. Rodd, F. H. & Reznick, D. N. Life-history evolution in guppies: III. The impact of prawn predation on guppy life histories. *Oikos* **62**, 13–19 (1991).
23. Reznick, D. N., Butler, M. J. 4th, Rodd, F. H. & Ross, P. Life-history evolution in guppies (*Poecilia reticulata*): 6. Differential mortality as a mechanism for natural selection. *Evolution* **50**, 1651–1660 (1996).
24. Bryant, M. J. & Reznick, D. Comparative studies of senescence in natural populations of guppies. *Am. Nat.* **163**, 55–68 (2004).
25. Endler, J. A. A predator's view of animal color patterns. *Evol. Biol.* **11**, 319–364 (1978).
26. Farr, J. A. Social behaviour patterns as determinants of reproductive success in the guppy, *Poecilia reticulata* Peters (Pisces, Poeciliidae). An experimental study of the effects of intermale competition, female choice, and sexual selection. *Behaviour* **74**, 38–91 (1980).
27. Bond, A. B. & Kamil, A. C. Visual predators select for crypticity and polymorphism in virtual prey. *Nature* **415**, 609–613 (2002).
28. Punzalan, D., Rodd, F. H. & Hughes, K. A. Perceptual processes and the maintenance of polymorphism through frequency-dependent predation. *Evol. Ecol.* **19**, 303–320 (2005).
29. Littell, R. C., Stroup, W. W. & Freund, R. J. *SAS for Linear Models* (SAS Institute, Cary, North Carolina, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Baril, M. Bryant, G. Cleven, K. Dixon, T. Hibler, A. Inman and T. Pitcher for help with field work; M. Bryant for advice on field experiments; Z. Columer, G. Chappel and K. Van Meter for help with scoring survival; and J. Burns, K. Dixon, M. Fitzpatrick, J. Leips, A. Price and C. Weadick for comments on the manuscript. This work was supported by grants from the National Science Foundation (to K.A.H. and A.E.H., and to D.N.R.) and by an NSERC grant to F.H.R. During 1996, K.A.H. was supported by an NIH NRSA Fellowship, and F.H.R. by the Center for Population Biology at the University of California at Davis. We thank the government of Trinidad and Tobago, and the Water and Sewage Authority, for permission to collect fish and conduct research.

Author Contributions All authors collected field data. K.A.H., F.H.R. and A.E.H. designed the experiment. K.A.H. and F.H.R. supervised the field work. D.P. conducted reliability analysis; R.O. scored survival; and R.O. and K.A.H. analysed the data and wrote the manuscript. All authors discussed and commented on the manuscript, and suggested revisions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.A.H. (kahughes@life.uiuc.edu).

Assembly of the inner rod determines needle length in the type III secretion injectisome

Thomas C. Marlovits^{1,2†}, Tomoko Kubori^{1†}, María Lara-Tejero¹, Dennis Thomas³, Vinzenz M. Unger² & Jorge E. Galán¹

Assembly of multi-component supramolecular machines is fundamental to biology, yet in most cases, assembly pathways and their control are poorly understood. An example is the type III secretion machine, which mediates the transfer of bacterial virulence proteins into host cells¹. A central component of this nanomachine is the needle complex or injectisome, an organelle associated with the bacterial envelope that is composed of a multi-ring base, an inner rod, and a protruding needle². Assembly of this organelle proceeds in sequential steps that require the reprogramming of the secretion machine. Here we provide evidence that, in *Salmonella typhimurium*, completion of the assembly of the inner rod determines the size of the needle substructure. Assembly of the inner rod, which is regulated by the *InvJ* protein, triggers conformational changes on the cytoplasmic side of the injectisome, reprogramming the secretion apparatus to stop secretion of the needle protein.

Type III secretion systems (TTSSs) are present in many bacteria pathogenic for humans, animals or plants^{1–3}. An essential component of this system is the ‘needle complex’ or ‘injectisome’, which is composed of a multi-ring base and a protruding needle (Supplementary Fig. S1a). In *Salmonella* spp., the cause of food poisoning and typhoid fever, the base is composed of three proteins, namely *InvG*, *PrgH* and *PrgK* (ref. 4). The needle, which is composed of a single protein, *PrgI*, is connected to a different substructure, the inner rod, which traverses the entire length of the base and is made of *PrgJ* (ref. 5). The inner rod is anchored to a socket-like structure on the basal side of the base, which may serve as an adaptor between the helical inner rod and the N-fold symmetric base⁵. The entire needle complex is traversed by a channel that serves as a conduit for the passage of proteins that travel the type III secretion pathway. Assembly of the needle complex involves the formation of the base, which upon recruitment of several accessory proteins, proceeds to secrete exclusively the proteins that make up the needle and inner rod substructures⁶. Once the needle reaches a certain length, the TTSS switches substrate specificity and becomes competent for the secretion of effector proteins that are to be delivered into the host cell. In *Salmonella*, the length of the needle and the substrate switching of the secretion machine depends on the accessory protein *InvJ* (refs 4, 7). Consequently, strains lacking *InvJ* assemble injectisomes with much longer needles and cannot secrete effector proteins. The mechanisms by which *InvJ* carries out its function are poorly understood. A *Yersinia* spp. homologue of *InvJ*, *YscP*, has been proposed to work as a ‘molecular ruler’. *YscP* would function as an extended polypeptide anchored to the tip of the growing needle and the base of the needle complex⁸. Its full extension would signal that

the needle has reached the appropriate length, conveying information to the secretion apparatus to switch substrates.

To investigate the mechanism of needle-length control and substrate switching in TTSSs, we compared by electron cryomicroscopy the detailed structural organization of needle complexes obtained from wild-type and $\Delta invJ$ *S. typhimurium* mutant strains. As previously observed⁴, needle complexes obtained from the $\Delta invJ$ *S. typhimurium* strain exhibited very long needle filaments (Supplementary Fig. S1b). In sharp contrast with wild-type needle complexes, however, the needles of these injectisomes became easily detached upon isolation and were almost never seen attached to the base (Supplementary Fig. S1b). As previously reported for wild type⁵, the base of the injectisomes of the $\Delta invJ$ mutant strain exhibit heterogeneous symmetries ranging from 19- to 22-fold, with 20- and 21-fold corresponding to the most abundant species (data not shown). Images from complexes isolated from $\Delta invJ$ *S. typhimurium* were sorted in different symmetry classes, and the three-dimensional structure for the 20-fold needle complex was reconstructed at 20 Å resolution (Supplementary Figs S2 and S3). Overall, the surface rendering of the needle complex of $\Delta invJ$ *S. typhimurium* shows no significant difference with that of wild type, exhibiting similar disposition and appearance of the outer and inner rings and their subunits⁵ (Fig. 1).

The cropping of the reconstruction of the $\Delta invJ$ *S. typhimurium* needle complex, however, revealed very significant structural differences within the base. Most notably, all particles isolated from the

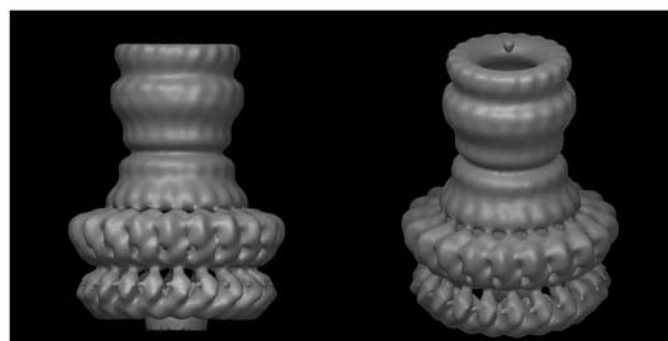


Figure 1 | Structure of TTSS injectisome obtained from $\Delta invJ$ *S. typhimurium*. Surface rendering of the structure of the $\Delta invJ$ *S. typhimurium* needle complex. Shown here is the structure of the 20-fold complex with 20-fold symmetry imposed (left panel, lateral view; right panel, tilted view).

¹Section of Microbial Pathogenesis, Yale University School of Medicine, Boyer Center for Molecular Medicine, New Haven, Connecticut 06536, USA. ²Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, Connecticut 06520-8024, USA. ³Howard Hughes Medical Institute, Rosenstiel Basic Medical Sciences Research Center, Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02454, USA. [†]Present addresses: Research Institute of Molecular Pathology (IMP) and Institute of Molecular Biotechnology, Austrian Academy of Sciences (IMBA), Dr Bohr-Gasse 5, A-1030 Vienna, Austria (T.C.M.); 21st Century COE Program, Combined Program on Microbiology and Immunology, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871, Japan (T.K.).

$\Delta invJ$ mutant lacked the inner rod substructure that traverses the base (Fig. 2a). Consistent with this observation, PrgJ, the main component of the inner rod⁵, was not detected in needle complexes obtained from the $\Delta invJ$ mutant strain (Fig. 2b) although this protein was readily detected in culture supernatants free from the shed needles^{4,9}. In addition, most of the $\Delta invJ$ particles lacked the characteristic 'socket' present on the basal plate of the base chamber of the wild-type injectisomes (Fig. 2a). The difference between the density maps of bases obtained from wild type and the $\Delta invJ$ mutant (Fig. 2c) indicates that the changes are indeed largely localized to the socket structure, so 'misalignments' between the two structures cannot account for the observed differences. As InvJ is not an integral component of the base^{2,4}, we hypothesize that this protein may help to stabilize the appropriate conformation of the socket that allows the 'docking' of PrgJ and the initiation (or termination) of the assembly of the inner rod. The failure of PrgJ to assemble into inner rods in the absence of InvJ, and its presence in culture supernatants of the $\Delta invJ$ mutant⁹, are consistent with this hypothesis.

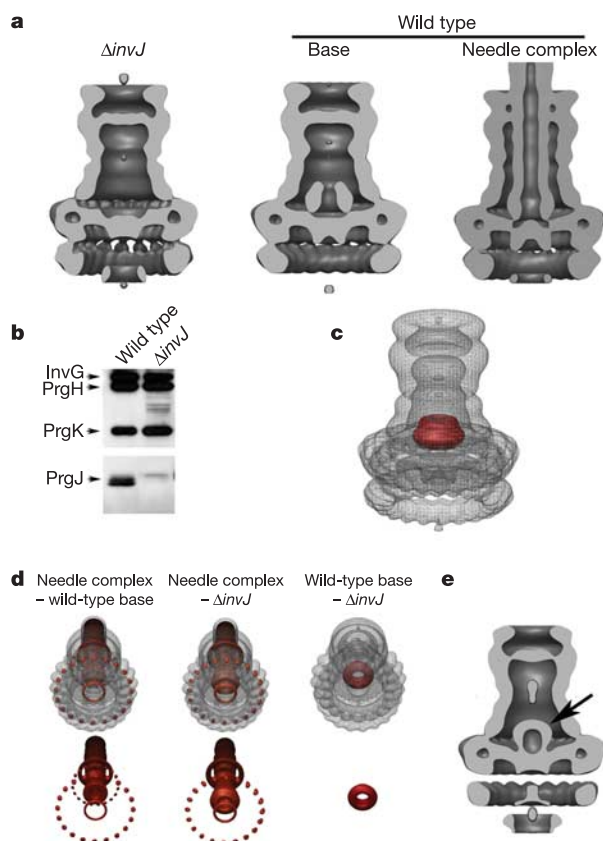


Figure 2 | Structural differences between needle complexes of wild-type, $\Delta invJ$ and $\Delta prgJ$ *S. typhimurium*. **a**, Cropping of the surface renderings of the $\Delta invJ$ injectisome base reveals the absence of the socket-like structure and the inner rod observed in the wild-type base and needle complexes, respectively. **b**, PrgJ, the main component of the inner rod, is absent from *S. typhimurium* $\Delta invJ$ needle complexes. Needle complexes obtained from wild-type and $\Delta invJ$ *S. typhimurium* strains were analysed by western immunoblotting. **c**, The difference between the density maps (indicated in red) of the wild-type (mesh) and the $\Delta invJ$ mutant base structures indicates that the socket is absent from the $\Delta invJ$ base. **d**, The density maps of the wild-type or $\Delta invJ$ base were subtracted from the maps of the fully assembled needle complex or the wild-type base, as indicated. The differences between the density maps are indicated in red, and the densities of the subtracted structures are indicated by mesh. The different structures are shown tilted forward 60°. **e**, Cropping of the surface rendering of the structure of the $\Delta prgJ$ injectisome base revealing the presence of the socket-like structure (arrow) that is observed in the wild-type base.

We have previously shown that the transition from the base to the fully assembled needle complex results in very significant conformational changes in the base substructure³. In addition to significant conformational changes observed within the base to accommodate the inner rod and needle, marked conformational changes were observed on the cytoplasmic face of the base substructure, which we hypothesized may provide the structural bases for substrate switching upon needle assembly⁵. Notably, a cup-like protrusion located on the cytoplasmic face of the base substructure, which may serve as entry point for proteins that travel this pathway, undergoes a significant 'downward' movement, therefore appearing longer in the fully assembled needle complex. Furthermore, completion of the needle complex results in significant conformational changes in one of the inner rings of the base, further redefining the shape of its cytoplasmic face³. The tilted views of the difference between the density maps of bases obtained from wild type and the $\Delta invJ$ mutant revealed that the cup-like protrusion on the cytoplasmic face is significantly shorter than that of the wild-type fully-assembled needle complexes (Fig. 2d). Additional conformational changes were also detected in one of the inner rings (Fig. 2d). Indeed, the cytoplasmic face of the $\Delta invJ$ base closely resembles that of the base substructure of the wild-type injectisome before needle assembly (Fig. 2d). This observation is consistent with the phenotype of the $\Delta invJ$ mutant, which is unable to undergo substrate switching and is therefore 'locked' in a secretion mode similar to that of the base before needle assembly—that is, competent for the secretion of the needle and inner rod proteins PrgI and PrgJ, but unable to secrete effector proteins^{4,6,7}.

We determined the three-dimensional structure of needle complexes obtained from a *S. typhimurium* $\Delta prgJ$ mutant. Although the resolution of this structure is lower than those of the $\Delta invJ$ and wild type (Supplementary Fig. S1), the cropping of this reconstruction

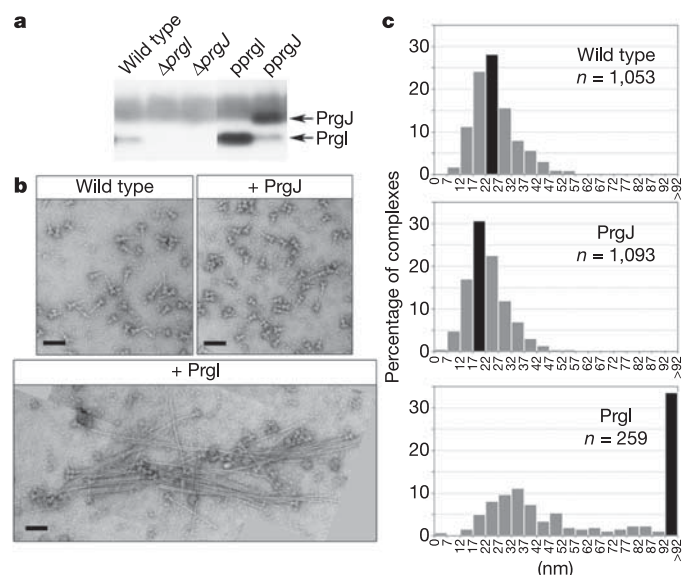


Figure 3 | Altering the stoichiometry between the needle and rod proteins results in injectisome needles of different lengths. **a**, Western immunoblot analysis of needle complexes from strains overexpressing the needle protein PrgI (labelled prgI) or the inner rod protein PrgJ (labelled prgJ). **b**, Electron micrographs of negatively stained needle complexes from *S. typhimurium* strains overexpressing either PrgI or PrgJ, as indicated. Scale bars, 100 nm. **c**, Measurements of needle length in needle complexes from wild-type *S. typhimurium* and derivative strains overexpressing either PrgI or PrgJ, as indicated. Dark bars denote the most abundant size of the measured species and *n* denotes the number of particles measured. The differences between the needle size of wild-type and PrgJ or PrgI overexpressing strains were statistically significant ($P < 0.0001$).

(Fig. 2e) clearly revealed the presence of the socket within the base, arguing that the stabilization of the socket and presumably the subsequent assembly of the inner rod is not the result of the direct interaction between InvJ with PrgJ. Rather, these results indirectly argue in favour of the hypothesis that InvJ exerts its function on the components of the base (that is, PrgH/PrgK) either directly or indirectly through interactions with other accessory proteins.

The observation that the $\Delta invJ$ mutant, which lacks the inner rod, produces needles indicates that needle assembly *per se* does not require the inner rod. However, the resulting needles in the $\Delta invJ$ mutant are loosely linked to the base and easily detach during purification (Supplementary Fig. S1b), indicating that assembly of the inner rod may be required for the proper anchoring of the needle substructure. Without the inner rod the type III secretion organelle is unable to secrete effector proteins and remains locked in a conformation that is only competent for the secretion of PrgI and PrgJ (refs 4, 6, 7). Therefore, termination of the assembly of the inner rod may be necessary to induce conformational changes that may be required for substrate switching, thereby controlling needle length. If this were the case, the needle length should be influenced by the relative concentration of the inner rod and needle proteins PrgJ and PrgI, which in turn may influence the speed at which each of the substructures is assembled.

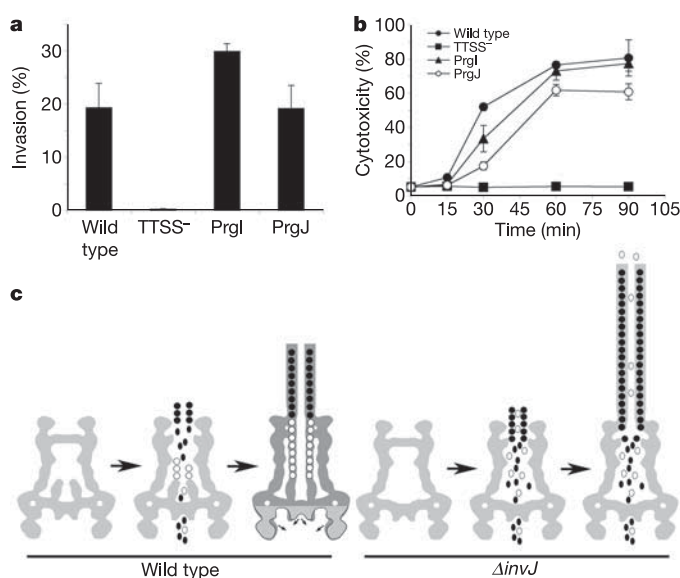


Figure 4 | Altering needle length by changes in the stoichiometry of the needle and inner rod proteins results in functional injectisomes. **a**, Levels of internalization into cultured epithelial cells of wild-type and TTSS-defective (TTSS⁻) *S. typhimurium* and derivative strains overexpressing either the needle protein PrgI or the inner rod protein PrgJ. Values are the means $\pm$ standard deviation of three independent measurements, and represent the percentage of the inoculum that survived treatment with gentamicin, an indication of bacterial internalization. **b**, Macrophage cytotoxicity of wild-type and TTSS-defective *S. typhimurium* and strain derivatives overexpressing either PrgI or PrgJ. Cytotoxicity was determined by assaying LDH release. Values represent the means $\pm$ standard deviation of three independent measurements and were standardized to 100% release obtained by complete cell lysis. **c**, Model of needle length control. Assembly of the inner rod but not the needle requires InvJ. Completion of the inner rod leads to conformational changes on the cytoplasmic side of the injectisome, which results in substrate switching and the interruption of secretion of the inner rod and needle proteins, hence determining the length of the needle substructure. In the absence of InvJ, the socket of the base is not formed and the inner rod fails to assemble. Consequently, the secretion machinery remains locked in a secretion mode competent for the secretion of the needle protein, resulting in non-functional injectisomes with abnormally long needles. Filled circles, needle protein PrgI; open circles, inner rod protein PrgJ.

To test this hypothesis, we constructed *S. typhimurium* strains that overexpress either PrgI or PrgJ (Fig. 3a) and examined the length of the needles in isolated needle complexes. Overexpression of PrgI resulted in injectisomes with needles significantly ($P < 0.0001$) longer than wild type (Fig. 3b). In fact, a significant proportion of the needles were too long to be accurately measured as they could not be observed in a single field of view. Unlike the long needles observed in the $\Delta invJ$ mutant strain (Fig. 1b), however, the vast majority of the long needles in the PrgI overexpressing strain were firmly attached to the base (Fig. 3b and c). Injectisomes obtained from a strain overexpressing PrgJ had needles that were significantly ($P < 0.0001$) shorter than wild type (Fig. 3b and c). Other than the different length of the needles, injectisomes obtained from these strains appeared identical to wild type. These results indicate that altering the stoichiometry of the inner rod and needle proteins has a profound impact on the length of the needle substructure.

To ascertain whether needle complexes with needles of different length assembled by *S. typhimurium* strains with an altered PrgI/PrgJ stoichiometry were functional, we examined two phenotypes that are known to depend on this TTSS: bacterial internalization into epithelial cells and cytotoxicity towards macrophages¹⁰. *S. typhimurium* overexpressing PrgI, which leads to longer needles, entered epithelial cells with slightly higher efficiency than wild type, although its cytotoxicity towards macrophages was not altered (Fig. 4a and b). This is in sharp contrast with the $\Delta invJ$ mutant, which exhibits longer needles but has a non-functional TTSS and is unable to enter epithelial cells^{7,11}. A strain overexpressing PrgJ, which leads to shorter needles, showed a slight decrease in macrophage cytotoxicity although it was not altered in its ability to enter epithelial cells (Fig. 4a and b). These results indicate that needles of different length caused by changes in the stoichiometry of PrgI and PrgJ are indeed functional.

Taken together, our results suggest a model for the mechanisms of needle-length control (Fig. 4c). Completion of the inner rod would result in the firm anchoring of the needle to the base, triggering conformational changes on the cytoplasmic side of the injectisome and reprogramming of the secretion apparatus to stop secretion of the needle and inner rod proteins. In this model, InvJ is required to stabilize a conformation of the socket of the injectisome base that is permissive for the anchoring of the inner rod. In the absence of InvJ, the socket is not formed, the inner rod fails to assemble or anchor, and the secretion machinery remains locked in a secretion mode that is competent for the secretion of the needle protein but not the effectors, resulting in non-functional injectisomes with abnormally long needles. The function of InvJ would be similar to that proposed for scaffolding proteins involved in flagellar assembly, which, like InvJ, do not form part of the final flagellar structure and are secreted to the culture medium^{12,13}. InvJ may therefore function as a 'kinetic trap', allowing the socket to go from a flexible, unstructured state to a state competent for interactions with the inner rod protein PrgJ. Consistent with this hypothesis, conformational heterogeneity in the 'socket' region of the needle complex can be detected when lowering the threshold of the reconstruction of the $\Delta invJ$ structure. Although TTSSs are largely conserved, more experiments will be required to extend the generality of this proposed mechanism to related systems in other bacteria.

METHODS

Bacterial strains and plasmids. All *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) were derived from strain SJW2941 and have been previously described⁶. Plasmids expressing either the PrgI or PrgJ protein were constructed by standard recombinant DNA techniques using the plasmid expression vector pWSK29 (ref. 14).

Bacterial invasion and macrophage cytotoxicity assays. The ability of the different *S. typhimurium* strains to invade cultured intestinal epithelial cells or to kill cultured J774 macrophages was assayed by the gentamicin resistance and LDH (lactic dehydrogenase) release assays, respectively, as previously described^{15,16}.

Needle complex preparation and analysis, electron microscopy, and image processing. Needle complex sample preparation for electron microscopy, biochemical analysis, electron microscopy, image processing and the determination of the rotational symmetry was carried out as previously described⁵ with some modifications (see Supplementary Methods).

Determination of needle length in needle complexes. Needle complex preparations from the different bacterial strains grown to late stationary phase were prepared as described elsewhere⁵. To avoid size bias during isolation, all caesium chloride fractions were pooled before measurement. Samples were placed on glow-discharged Cu/Rh grids and negatively stained using 2% PTA (phospho-tungstic acid pH 7.2). Electron microscopic images were recorded with a 1K CCD camera using a Technai T12 electron microscope equipped with a LaB6-filament operated at 120 kV and 30,000 $\times$ magnification. Each pixel corresponded to a sampling of 5.5 Å at the level of the specimen. To determine the length of individual needle filaments protruding from the base, three data points were recorded: (1) the centre of inner ring 1, (2) the centre of outer ring 1, and (3) the tip of the needle filament (see Supplementary Fig. S1a). To account for tilted particles in the recordings, the length of the filaments (that is, the distance between (2) and (3)) was normalized to the constant distance between IR1 and OR1 (distance between (1) and (2)). Statistical analysis was carried out using the Mann–Whitney test.

Received 3 April; accepted 19 April 2006.

- Galán, J. E. & Collmer, A. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science* **284**, 1322–1328 (1999).
- Kubori, T. *et al.* Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**, 602–605 (1998).
- Cornelis, G. R. & Van Gijsegem, F. Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.* **54**, 735–774 (2000).
- Kubori, T., Sukhan, A., Aizawa, S. I. & Galán, J. E. Molecular characterization and assembly of the needle complex of the *Salmonella typhimurium* type III protein secretion system. *Proc. Natl Acad. Sci. USA* **97**, 10225–10230 (2000).
- Marlovits, T. C. *et al.* Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, 1040–1042 (2004).
- Sukhan, A., Kubori, T., Wilson, J. & Galán, J. E. Genetic analysis of assembly of the *Salmonella enterica* serovar Typhimurium type III secretion-associated needle complex. *J. Bacteriol.* **183**, 1159–1167 (2001).
- Collazo, C. & Galán, J. E. Requirement of exported proteins for secretion through the invasion-associated Type III system in *Salmonella typhimurium*. *Infect. Immun.* **64**, 3524–3531 (1996).
- Journet, L., Agrain, C., Broz, P. & Cornelis, G. R. The needle length of bacterial injectisomes is determined by a molecular ruler. *Science* **302**, 1757–1760 (2003).
- Sukhan, A., Kubori, T. & Galán, J. E. Synthesis and localization of the *Salmonella* SPI-1 type III secretion needle complex proteins PrgI and PrgJ. *J. Bacteriol.* **185**, 3480–3483 (2003).
- Galán, J. E. *Salmonella* interaction with host cells: Type III secretion at work. *Annu. Rev. Cell Dev. Biol.* **17**, 53–86 (2001).
- Collazo, C. M., Zierler, M. K. & Galán, J. E. Functional analysis of the *Salmonella typhimurium* invasion genes *invI* and *invJ* and identification of a target of the protein secretion apparatus encoded in the *inv* locus. *Mol. Microbiol.* **15**, 25–38 (1995).
- Macnab, R. How bacteria assemble flagella. *Annu. Rev. Microbiol.* **57**, 77–100 (2003).
- Macnab, R. Type III flagellar protein export and flagellar assembly. *Biochim. Biophys. Acta* **1694**, 207–217 (2004).
- Wang, R. F. & Kushner, S. R. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. *Gene* **100**, 195–199 (1991).
- Chen, L. M., Kaniga, K. & Galán, J. E. *Salmonella* spp. are cytotoxic for cultured macrophages. *Mol. Microbiol.* **21**, 1101–1115 (1996).
- Galán, J. E. & Curtiss, R. III Cloning and molecular characterization of genes whose products allow *Salmonella typhimurium* to penetrate tissue culture cells. *Proc. Natl Acad. Sci. USA* **86**, 6383–6387 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Galán laboratory for critical reviews of this manuscript. We are also grateful to the Yale School of Medicine for the support of the Cryo Electron Microscopy Core Facility. Molecular graphic images were produced using the Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by the NIH). This work was supported by grants from the National Institutes of Health (to V.M.U. and J.E.G.).

Author Information The three-dimensional density maps have been deposited into the EBI-MSD EMD database under the accession code EMD-1214. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.E.G. (jorge.galan@yale.edu).

The inner-nuclear-envelope protein emerin regulates HIV-1 infectivity

Jean-Marc Jacque^{1†} & Mario Stevenson¹

Primate lentiviruses such as human immunodeficiency type 1 (HIV-1) have the capacity to infect non-dividing cells such as tissue macrophages¹. In the process, viral complementary DNA traverses the nuclear envelope to integrate within chromatin². Given the intimate association between chromatin and the nuclear envelope³, we examined whether HIV-1 appropriates nuclear envelope components during infection. Here we show that emerin, an integral inner-nuclear-envelope protein, is necessary for HIV-1 infection. Infection of primary macrophages lacking emerin was abortive in that viral cDNA localized to the nucleus but integration into chromatin was inefficient, and conversion of viral cDNA to non-functional episomal cDNA increased. HIV-1 cDNA associated with emerin *in vivo*, and the interaction of viral cDNA with chromatin was dependent on emerin. Barrier-to-autointegration factor (BAF), the LEM (LAP, emerin, MAN) binding partner of emerin, was required for the association of viral cDNA with emerin and for the ability of emerin to support virus infection. Therefore emerin, which bridges the interface between the inner nuclear envelope and chromatin, may be necessary for chromatin engagement by viral cDNA before integration.

Components of the nuclear envelope include the inner and outer nuclear membranes, nuclear pore complexes and the nuclear lamina (reviewed in ref. 4). Lamins associate with several proteins including nesprins and integral nuclear envelope LEM-domain proteins, which have functions in chromatin organization and gene expression as well as in signal transduction³. We used primary macrophages as the experimental system in which to examine the requirement for components of the nuclear envelope in HIV-1 infection. Macrophages are highly relevant for the study of HIV-1 and, because they are non-dividing, the role of the nuclear envelope in HIV-1 infection can be studied without resorting to the use of cell-cycle inhibitors. In addition, primary macrophages are highly amenable to short interfering RNA (siRNA)-mediated gene silencing^{5,6}. The susceptibility of primary macrophages to HIV-1 infection was examined after siRNA-mediated silencing of nuclear lamins and several lamin-associated proteins (Fig. 1a). Previous biochemical approaches implicated high-mobility-group protein A1 (HMGA1) as an HIV-1 integration cofactor^{7,8}. Gene knockout experiments subsequently revealed HMGA1 as nonessential in dividing cells⁹ and in multiple experiments with macrophages from different donors, HMGA1 was dispensable for infection in primary non-dividing cells, as was the

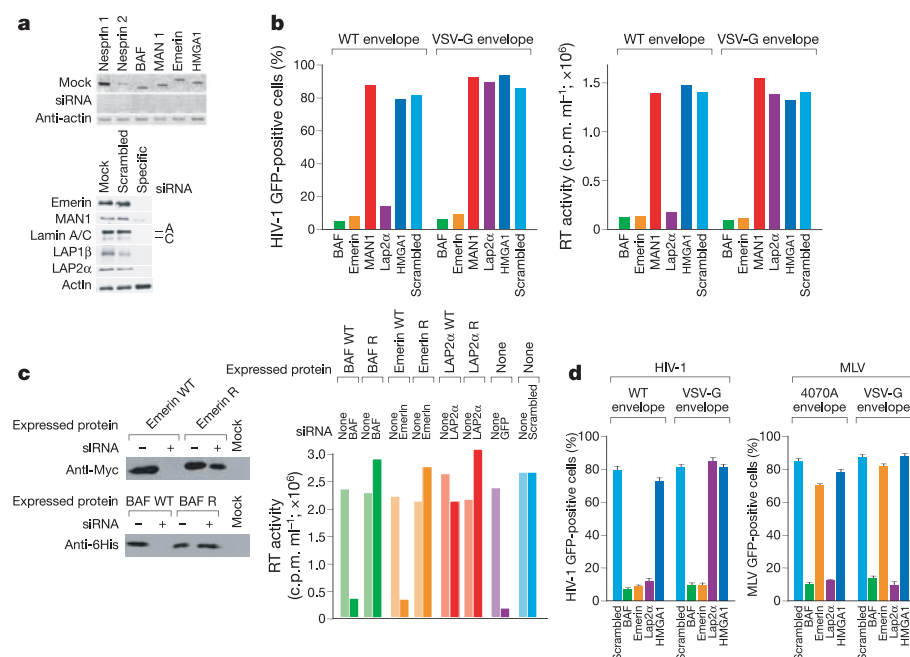


Figure 1 | siRNA-mediated silencing of nuclear envelope proteins inhibits HIV-1 infection of primary macrophages. **a**, siRNA-mediated silencing of nuclear envelope expression in primary macrophages. The extent of silencing was gauged by RT-PCR (top) and western blotting (bottom). **b**, The impact of silencing on the efficiency of macrophage infection by VSV-G-pseudotyped (VSV-G envelope) and non-pseudotyped (WT envelope) HIV-1 variants that express GFP in place of the viral NEF protein was gauged from the efficiency of GFP-positive cells (left) and from levels of viral reverse transcriptase released into culture supernatants (right). The standard deviation was less than $\pm 2\%$ ($n = 3$). **c**, Expression of ectopic wild-type (WT) proteins but not siRNA-resistant proteins (R) is inhibited by emerin and BAF siRNAs as shown by western blotting (left). The expression of siRNA-resistant forms of emerin and BAF in macrophages before silencing of endogenous proteins renders them fully susceptible to infection with VSV-G-pseudotyped HIV-1 (right). **d**, Effect of nuclear envelope silencing on infection of dividing HeLa cells. Cells were infected with VSV-G-pseudotyped and non-pseudotyped HIV-1 and amphotrophic (4070A) and VSV-G-pseudotyped MLV variants expressing a GFP gene. Error bars are s.d.

¹Program in Molecular Medicine, University of Massachusetts Medical School, 373 Plantation Street, Suite 319, Worcester, Massachusetts 01605, USA. [†]Present address: Centre for Research in Infectious Diseases, University College Dublin, Belfield, Dublin 4, Ireland.

LEM-domain protein MAN1 (Fig. 1b). Nesprin 1 and 2, lamin A/C, lamin B1 and B2 and LAP1 β were also dispensable for infection (not shown). In contrast, infection by vesicular stomatitis virus (VSV)-G-pseudotyped and non-pseudotyped HIV-1 was inefficient in macrophages in which expression of the LEM protein emerlin was inhibited (Fig. 1b). The LEM protein LAP2 α was necessary for infection by HIV-1 harboring a wild-type envelope but not for infection by VSV-G envelope-pseudotyped HIV-1 (Fig. 1b). BAF, which binds directly to the LEM domains of MAN1, LAP and emerlin, was also required for the infection of macrophages by both pseudotyped and non-pseudotype HIV-1 (Fig. 1b). The infectivity defect was not due to off-target effects on cellular activities unrelated to emerlin or BAF silencing, because expression of siRNA-resistant forms of emerlin or BAF restored the susceptibility of macrophages to HIV-1 infection (Fig. 1c). Murine leukaemia virus (MLV) infection, which is restricted to dividing cells¹⁰, has been shown to require the LEM protein LAP2 α (ref. 11). We therefore compared the effects of emerlin and BAF silencing on the infection of dividing HeLa cells by MLV and HIV-1 (Fig. 1d). Infection of dividing HeLa cells by pseudotyped and non-pseudotyped HIV-1 was dependent on emerlin and BAF (Fig. 1d). By comparison, infection of these cells by amphotropic

and VSV-G-pseudotyped MLV was dependent on LAP2 α and BAF but independent of emerlin (Fig. 1d). Because emerlin-depleted cells supported efficient MLV infection, this result excludes the possibility that the effects of emerlin silencing on HIV-1 infection are due to global effects on cell physiology. In addition, macrophages depleted of emerlin and BAF were fully susceptible to infection by an adenovirus vector expressing green fluorescent protein (GFP)¹², and transduction efficiencies for BAF-depleted, emerlin-depleted and mock-depleted macrophages were 92–95%. Through interaction with HIV-1 gag p55 and p17, BAF is encapsidated in virions at low levels¹³. However, virion-associated BAF did not promote HIV-1 infection because viruses produced by HeLa cells lacking BAF or emerlin were fully infectious (Supplementary Fig. S1). Collectively, these experiments demonstrate a requirement for emerlin and its LEM binding partner BAF in HIV-1 infection. In addition, HIV-1 and MLV appropriate different inner-nuclear-envelope proteins for infection.

After infection of a cell, linear viral cDNAs, generated by reverse transcription, translocate to the nucleus, where they integrate within cellular DNA to form a provirus. Within the nucleus, a certain proportion of linear cDNAs are inactivated by cellular DNA-

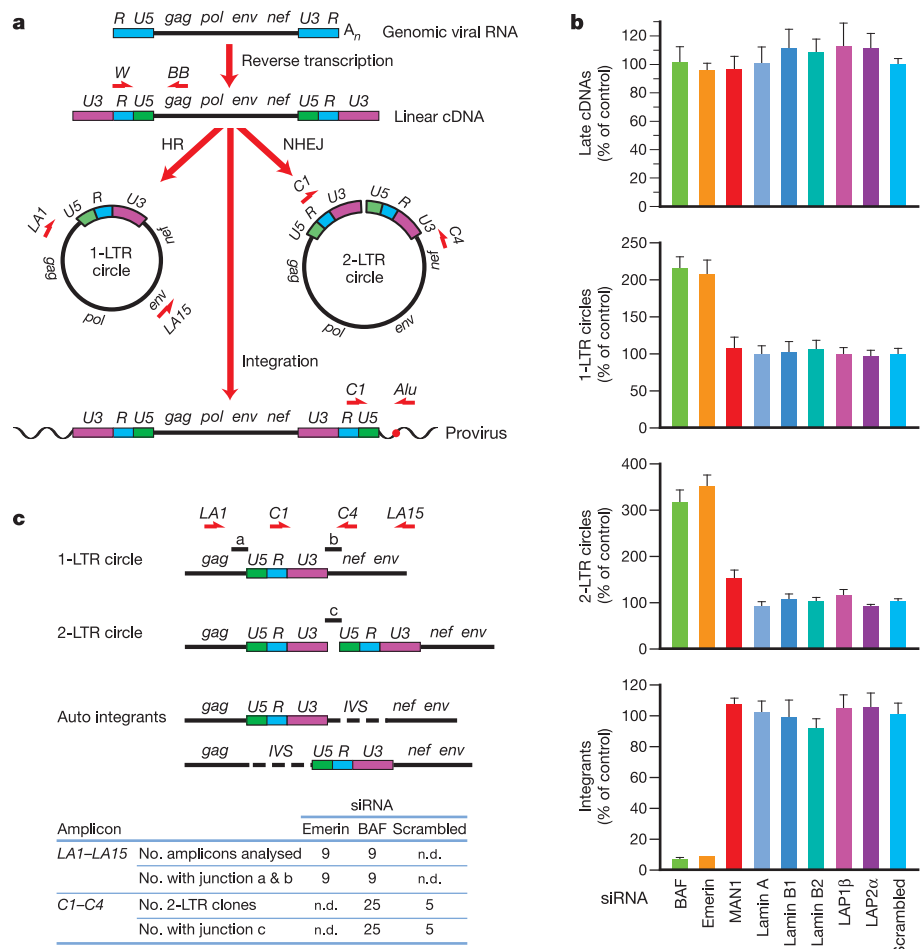


Figure 2 | Emerin and BAF silencing promotes abortive infection of macrophages by preventing viral cDNA integration. **a**, Strategy for analysis of viral cDNA. Linear cDNA integrates into chromatin to form a provirus; this is essential for productive infection of a cell. Some linear cDNAs are converted by homologous recombination (HR) and non-homologous end-joining (NHEJ) to form 1-LTR and 2-LTR circles, respectively. **b**, After transfection of macrophages with the indicated siRNAs, macrophages were infected (VSV-G-pseudotyped HIV-1) and levels of viral reverse transcription products were determined. cDNA copy number, calculated on the basis of CCR5 gene copies, was expressed as a percentage of scrambled

siRNA-transfected cells. The GFP siRNA inhibits HIV-1 infection by degrading incoming genomic viral RNA as well as viral transcripts generated *de novo* after integration²⁷. Error bars are s.d. **c**, Lack of autointegration after infection of macrophages lacking emerlin or BAF. Top: autointegration, which arises from intramolecular integration, leads to juxtaposition of LTR sequences and random intervening viral sequences (IVS). Bottom: amplicons generated by LA1–LA15 or C1–C4 primers exclusively harboured 1-LTR junction sequences (junctions a and b) or 2-LTR junction sequences (junction c), respectively. n.d., not done.

strand-break (DSB) repair enzymes. In this process, linear viral cDNAs are converted by homologous recombination or end–end ligation into circular cDNA forms containing one or two long terminal repeats (1-LTR or 2-LTR circles, respectively; Fig. 2a). Because these episomal cDNAs are unable to integrate, they are abortive products of infection. Viral cDNA synthesis (late cDNAs) in infected macrophages lacking emerlin or BAF was comparable to that in macrophages transfected with scrambled siRNAs (Fig. 2b). In contrast, provirus establishment was markedly impaired and there was a corresponding increase in the formation of 1-LTR and 2-LTR circles in these cells (Fig. 2b). In this experiment, episomal cDNAs constituted a minor fraction of the total cDNA pool in cells depleted of emerlin and BAF (less than 20% of late cDNA). Nevertheless, integration was still markedly impaired. Increased episomal cDNA formation could therefore not explain the defect in viral integration in the absence of emerlin or BAF. BAF was first identified from its ability to protect MLV cDNA from autointegration *in vitro*^{14,15}. Autointegration results in aberrant episomal junction sequences in which viral LTR sequences are randomly juxtaposed at intervening viral sequences (Fig. 2c). As a consequence, autointegrants do not harbour gag–U5 and U3–nef junctions characteristic of 1-LTR circles, nor U3–U5 junctions contained within 2-LTR circles (Fig. 2c). Episomal cDNAs cloned from macrophages lacking BAF

or emerlin had normal LTR junction sequences and were therefore a result of end–end ligation and homologous recombination (Fig. 2c). The impairment of integration in cells depleted of BAF and emerlin therefore occurred without a noticeable increase in the frequency of autointegration.

To obtain further information on the nature of the block to HIV-1 integration in the absence of emerlin and BAF, we examined the subnuclear distribution¹⁶ of viral cDNAs. In control macrophages transfected with scrambled or MAN1 siRNAs, most viral cDNA was sequestered in a chromatin-enriched cell fraction (Fig. 3a). In contrast, in macrophages lacking emerlin or BAF, viral cDNAs inefficiently associated with chromatin, whereas most were sequestered in the nuclear matrix (Fig. 3a). This was also reflected by a marked decrease in the number of proviruses within the chromatin fraction and a greater proportion of episomal cDNAs in the nuclear matrix (Fig. 3a). To examine whether the decreased association of viral cDNA with chromatin was a result of defects in integration, the subnuclear distribution of viral cDNA was examined after infection with an HIV-1 integrase catalytic site mutant (HIV-1 IN_(N,N))¹⁷. This variant associated with viral cDNA but failed to integrate (Fig. 3b). When emerlin and BAF were silenced, association of the integrase catalytic mutant with chromatin was also impaired (Fig. 3b). Mislocalization of viral cDNA in the absence of emerlin and BAF was

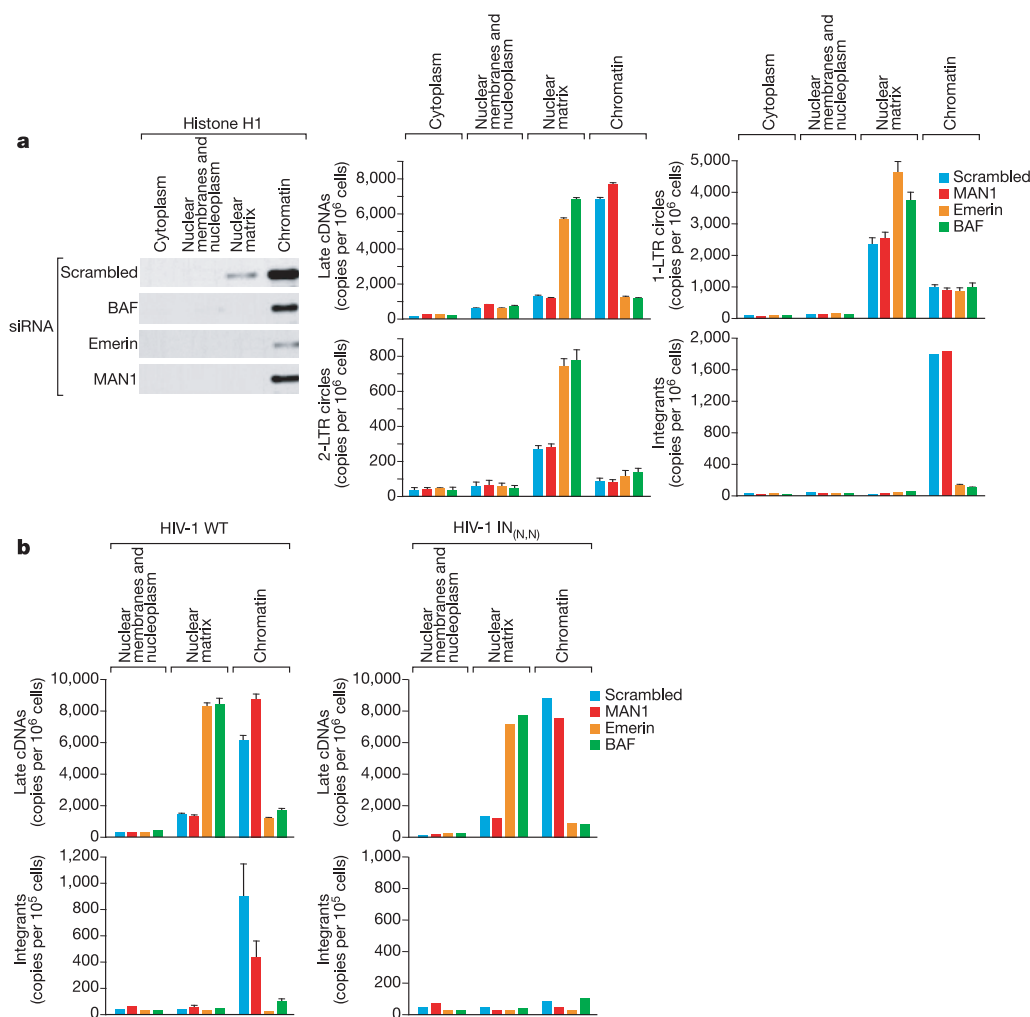


Figure 3 | Association of HIV-1 cDNA with chromatin requires BAF and emerlin. **a**, Subnuclear distribution of viral cDNAs after infection of macrophages depleted of emerlin or BAF with wild-type HIV-1. Top: western blot analysis of chromatin-associated protein histone H1 in subcellular fractions. The relative distributions of viral cDNA products after siRNA-

mediated silencing of nuclear envelope proteins are shown at the right. **b**, Subnuclear distribution of viral cDNAs after infection with an HIV-1 variant (HIV-1 IN_(N,N)) harbouring a catalytically inactive integrase domain. Error bars are s.d.

therefore unlikely to be a consequence of impaired integration. In the absence of emerin or BAF, most episomal viral cDNA was in the form of 1-LTR circles. To determine whether emerin and BAF increased the efficiency of productive HIV-1 infection by counteracting the action of the HR repair pathway, we examined the susceptibility of macrophages to HIV-1 infection after simultaneous silencing of emerin or BAF and the HR repair protein RAD52, which promotes 1-LTR circle formation¹⁸. There was a specific decrease in the formation of 1-LTR circles in macrophages lacking RAD52 (Supplementary Fig. S2). Although simultaneous silencing of RAD52 with either emerin or BAF also led to a marked decrease in 1-LTR circle formation, this did not negate the defect imposed by the silencing of either emerin or BAF (Supplementary Fig. S2). This indicates that the infectivity defect imposed by the silencing of emerin or BAF might be independent of the extent to which viral cDNA is converted to episomal cDNAs.

Independent silencing of BAF or emerin induced identical infectivity defects (Figs 2 and 3), indicating a possible cooperative role for these proteins in HIV-1 infection. We used a modified chromatin immunoprecipitation (ChIP) approach to examine whether emerin and BAF associate cooperatively with HIV-1 cDNA *in vivo*. Macrophages were transduced with lentivirus vectors expressing His-tagged BAF protein or Myc-tagged emerin. After macrophage infection, emerin and BAF proteins were immunoprecipitated with tag-specific antibodies and the presence of viral cDNA (late cDNAs) in these immunoprecipitates was determined by polymerase chain reaction (PCR). After acute infection of macrophages, viral cDNA co-immunoprecipitated with BAF and with emerin (Fig. 4a). The association of emerin with viral cDNA was abolished when BAF expression was silenced. However, interaction of BAF with viral cDNA was not affected when emerin expression was inhibited (Fig. 4a). BAF co-immunoprecipitates with HIV-1 preintegration complexes isolated from acutely infected cells, and the viral integrase protein is reported to mediate this association¹⁹. To explore the role of the viral integrase in the association, we repeated the analysis with HIV-1 variants containing a deletion in integrase (Δ IN)²⁰ or a Q168A mutation in integrase that compromises viral infectivity but not integration activity *in vitro*²¹. The presence of viral cDNA in cell lysates and in immunoprecipitates with tag-specific antibodies was then determined by real-time PCR to accommodate potential differences in the efficiency of reverse transcription by wild-type and integrase mutant viruses. Although levels of viral cDNA synthesis in wild-type and integrase-mutant-infected macrophages was nearly equivalent (Fig. 4b), viral cDNA association with emerin and BAF was evident in cells infected with the Q168A integrase mutant but not in cells infected with a mutant lacking HIV-1 integrase (Fig. 4b). Integrase is therefore required for the association of viral cDNA with BAF. To examine whether BAF supports HIV-1 infection only in the presence of emerin, we characterized determinants in BAF and in emerin that were required for viral cDNA association and for viral infectivity. Macrophages were transduced with lentivirus vectors expressing wild-type BAF or emerin, BAF mutants defective in emerin binding (BAF 25Q and 53E)²² or an emerin mutant lacking an LEM domain (emerin Δ LEM) that mediates interaction with BAF²³ (Fig. 4c). Viral cDNA did not associate with either the BAF 25Q or 53E mutants nor the emerin LEM domain mutant (Fig. 4c). On infection of macrophages expressing these emerin and BAF mutants, viral cDNA synthesis was not affected but the integration of viral cDNA was markedly impaired and there was a corresponding increase in episomal cDNA formation, namely a phenotype identical to that obtained in macrophages lacking emerin or BAF (Fig. 4d). Taking these results together, this demonstrates that emerin and BAF are cooperative cofactors of HIV-1 infection.

Our study provides evidence that HIV-1 cDNA, on entering the nucleus, must interact with emerin to contact chromatin. Our observation that emerin silencing impaired the infection of dividing cells indicates that, during infection of a dividing cell, emerin might

also promote chromatin engagement on reformation of the nuclear envelope at the end of mitosis. In support of this, MLV infection, which is restricted to dividing cells, has been demonstrated to be dependent on the nuclear envelope protein LAP2 α (ref. 11). Our study confirms that result, but we show further that MLV infection is independent of emerin. Although we found that LAP2 α was also required for HIV-1 infection, this requirement was bypassed during infection by VSV-pseudotyped HIV-1. Emerin and LAP2 α did not seem to have redundant functions in HIV-1 infection because the silencing of one was not compensated for by the presence of the other. It is possible that LAP2 α is required for a step in infection that is common to lentiviruses and onco-retroviruses, whereas emerin is a specific lentivirus cofactor. It remains to be determined whether the differences in the ability of HIV-1 and MLV to infect non-dividing cells are dictated by the specific nuclear envelope proteins that these viruses appropriate for infection.

The infectivity defect (decreased provirus formation and increased formation of episomes) imparted by the silencing of emerin and BAF is similar to that observed in cells infected with HIV-1 integrase

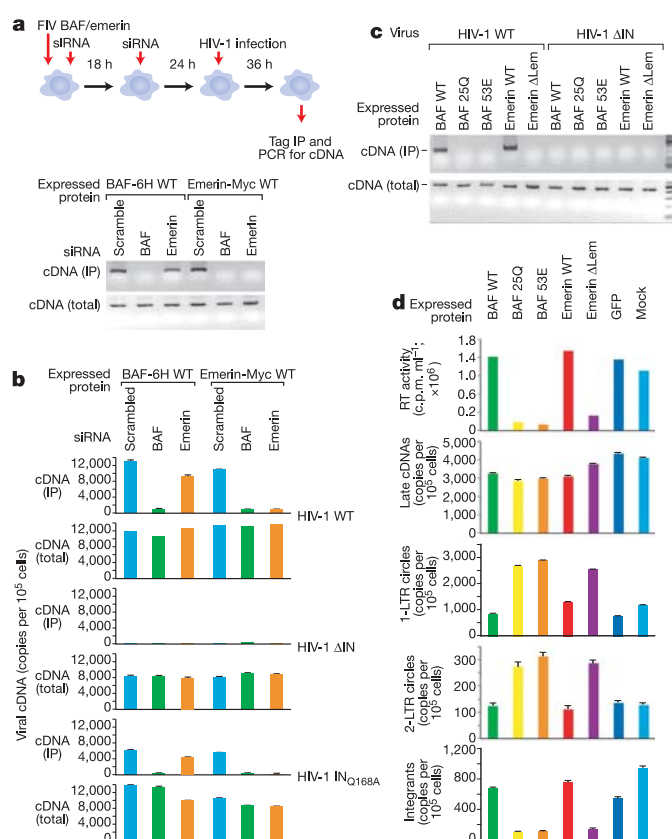


Figure 4 | HIV-1 cDNA association with emerin *in vivo* is dependent on BAF and requires the viral integrase. **a**, Top: experimental design. After FIV transduction and transfection with the indicated siRNAs, cell lysates were immunoprecipitated with tag-specific antibodies. Bottom: the presence of viral cDNA in cell lysates (total) and in immunoprecipitates (IP) was determined by PCR. siRNAs directed to BAF and emerin recognize both endogenous and expressed forms of these proteins. **b**, Deletion of the viral integrase domain impairs association of viral cDNA with either BAF or emerin. Viral cDNA copy number in total cell lysates and in immunoprecipitates was determined by real-time PCR. HIV-1 Δ IN contains a deletion in the viral integrase. HIV-1 IN_{Q168A} harbours a functional integrase, yet is replication defective²¹. **c**, Mutations in BAF and emerin that impair their association with viral cDNA. Association of viral cDNA with wild-type (WT) and mutant BAF and emerin proteins was examined in HIV-1-infected macrophages, as in **a**. **d**, Dominant-negative effects of mutant emerin and BAF proteins on viral cDNA integration and infection efficiency. Error bars are s.d.

mutants or HIV-1 infections conducted in the presence of integrase inhibitors^{24,25}. However, emerin and BAF did not facilitate cDNA integration but, rather, were required for the appropriate localization of viral cDNA before chromatin engagement. This provides the rationale for the identification of small molecules that antagonize the ability of the virus to engage emerin, because such agents would be predicted to promote abortive HIV-1 infection of a cell. In the X-linked form of Emery–Dreifuss muscular dystrophy, which is caused by mutations in the emerin gene, there is an almost complete absence of the protein in tissues²⁶. Further investigation is needed into whether sufferers of X-linked Emery–Dreifuss muscular dystrophy are resistant to HIV-1 infection or exhibit a non-progressive disease pattern on infection.

METHODS

Lentivirus vectors. The FIV packaging system was obtained from Systembio. The FIV-CopGFP vector was modified by replacing the GFP open reading frame by a stuffer containing an *SrfI* site. BAF and emerin were amplified by PCR with reverse transcription (RT–PCR) from 3 µg of total macrophage RNA (one-step RT–PCR; Invitrogen) and cloned into the FIV shuttle vector. siRNA-resistant forms of BAF, emerin and LAP2α were generated by a two-step PCR with internal primers 5′-GGCCTACGTGGTCTTGGC-3′, 5′-GAGGAGTGTAAGACAGGG-3′ and 5′-GACAAGTTAAAAAGCGAGTTGG-3′, respectively (silent point mutations are shown in bold). The BAF single-amino-acid substitutions²³ were similarly generated by replacing Gly25 with Gln (GGG to CAA) and Lys53 with Glu (AAG to GAG). The LEM mutant of emerin was generated by deleting amino-acid residues 2–44.

Proximal DNAs. VSV-G-pseudotyped HIV-1 were prepared by co-transfecting HeLa cells with pNL-GFP and pLVSVG²⁷. Macrophage-tropic-viruses were prepared by transfecting HeLa cells with pADA-GFP (An *EcoRI*–*Bam*HI fragment from pADAwT was inserted into pNLGFP). HIVΔIN contains a deletion in integrase²⁰. Amphotropic MLV and VSV-G-pseudotyped MLV was prepared by transfecting HeLa cells and plasmids expressing an amphotropic MLV (4070A) envelope and a VSV envelope, respectively. The HIV-1 integrase point mutants HIV-1 IN_{Q168A} and HIV-1 IN_(N,N) are as described²¹.

Subnuclear fractionation. Extractions were performed as described¹⁶, with minor modifications. In brief, cells were quickly lysed in 300 µl of buffer Tx-CSK (20 mM KPO₄ pH 6.8, 100 mM NaCl, 1 mM MgCl₂, 1 mM EGTA, 0.1% Triton X-100) and layered on 200 µl of sucrose solution (30% sucrose, 50 mM Tris-HCl pH 8.2, 5 mM MgCl₂, 0.1 mM EDTA). Intact nuclei were pelleted at 3,500 g for 5 min. Nuclear membranes and soluble proteins were dissolved in buffer N (40 mM Tris-HCl pH 6.8, 5 mM MgCl₂, 1.5 M NaCl, 0.5% Triton X-100). Nuclear matrix and chromatin were pelleted at 12,000 g for 20 min. Nuclear matrix proteins were subsequently extracted in buffer M (25 mM Tris-HCl pH 10, 0.5 M NaCl, 5 mM dithiothreitol, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100). Chromatin structures were pelleted at 14,000 g for 30 min and resuspended in TE buffer.

Protein–cDNA co-immunoprecipitations. Cells were lysed in triple-detergent lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% SDS, 1% Nonidet P40, 0.5% sodium deoxycholate, 10 µl ml^{−1} protease inhibitor cocktail (Sigma)) and incubated with anti-Myc antibody (Invitrogen) for emerin or Ni²⁺-nitrilotriacetate beads (Qiagen) (100 µl of solution pelleted and rinsed with lysis buffer) for BAF. The lysates were incubated for 2 h at 4 °C and washed five times in lysis buffer. Pellets were resuspended in 50 µl of TE buffer; 10 µl was used for PCR analysis.

Supplementary Methods contains details on siRNAs used for gene silencing, and on primers used for PCR analysis and for additional information on quantitative PCR analysis of viral cDNAs, RNA-mediated interference analysis in macrophages and HeLa cells, western blotting, and cell-cycle analysis.

Received 26 January; accepted 28 February 2006.

Published online 7 May 2006.

1. Stevenson, M. HIV-1 pathogenesis. *Nature Med.* **9**, 853–860 (2003).
2. Greene, W. C. & Peterlin, B. M. Charting HIV's remarkable voyage through the cell: Basic science as a passport to future therapy. *Nature Med.* **8**, 673–680 (2002).
3. Gruenbaum, Y., Margalit, A., Goldman, R. D., Shumaker, D. K. & Wilson, K. L. The nuclear lamina comes of age. *Nature Rev. Mol. Cell Biol.* **6**, 21–31 (2005).

4. Zastrow, M. S., Vlcek, S. & Wilson, K. L. Proteins that bind A-type lamins: Integrating isolated clues. *J. Cell Sci.* **117**, 979–987 (2004).
5. Song, E. et al. Sustained small interfering RNA-mediated human immunodeficiency virus type 1 inhibition in primary macrophages. *J. Virol.* **77**, 7174–7181 (2003).
6. Zielske, S. P. & Stevenson, M. Importin 7 may be dispensable for human immunodeficiency virus type 1 and simian immunodeficiency virus infection of primary macrophages. *J. Virol.* **79**, 11541–11546 (2005).
7. Farnet, C. M. & Bushman, F. D. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes *in vitro*. *Cell* **88**, 483–492 (1997).
8. Li, L. et al. Retroviral cDNA integration: stimulation by HMG I family proteins. *J. Virol.* **74**, 10965–10974 (2000).
9. Beitzel, B. & Bushman, F. Construction and analysis of cells lacking the HMGA gene family. *Nucleic Acids Res.* **31**, 5025–5032 (2003).
10. Lewis, P. F. & Emerman, M. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J. Virol.* **68**, 510–516 (1994).
11. Suzuki, Y., Yang, H. & Craigie, R. LAP2α and BAF collaborate to organize the Moloney murine leukemia virus preintegration complex. *EMBO J.* **23**, 4670–4678 (2004).
12. Swinger, S. et al. HIV-1 Nef mediates lymphocyte chemotaxis and activation by infected macrophages. *Nature Med.* **5**, 997–1003 (1999).
13. Mansharamani, M. et al. Barrier-to-autointegration factor BAF binds p55 Gag and matrix and is a host component of human immunodeficiency virus type 1 virions. *J. Virol.* **77**, 13084–13092 (2003).
14. Lee, M. S. & Craigie, R. A previously unidentified host protein protects retroviral DNA from autointegration. *Proc. Natl Acad. Sci. USA* **95**, 1528–1533 (1998).
15. Segura-Totten, M. & Wilson, K. L. BAF: roles in chromatin, nuclear structure and retrovirus integration. *Trends Cell Biol.* **14**, 261–266 (2004).
16. Miccoli, L. et al. Selective interactions of human kin17 and RPA proteins with chromatin and the nuclear matrix in a DNA damage- and cell cycle-regulated manner. *Nucleic Acids Res.* **31**, 4162–4175 (2003).
17. Lu, R. et al. Class II integrase mutants with changes in putative nuclear localization signals are primarily blocked at a postnuclear entry step of human immunodeficiency virus type 1 replication. *J. Virol.* **78**, 12735–12746 (2004).
18. Lau, A., Kanaar, R., Jackson, S. P. & O'Connor, M. J. Suppression of retroviral infection by the RAD52 DNA repair protein. *EMBO J.* **23**, 3421–3429 (2004).
19. Lin, C. W. & Engelman, A. The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. *J. Virol.* **77**, 5030–5036 (2003).
20. Stevenson, M. et al. Integration is not necessary for expression of human immunodeficiency virus type 1 protein products. *J. Virol.* **64**, 2421–2425 (1990).
21. Cherepanov, P., Ambrosio, A. L., Rahman, S., Ellenberger, T. & Engelman, A. Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc. Natl Acad. Sci. USA* **102**, 17308–17313 (2005).
22. Segura-Totten, M., Kowalski, A. K., Craigie, R. & Wilson, K. L. Barrier-to-autointegration factor: Major roles in chromatin decondensation and nuclear assembly. *J. Cell Biol.* **158**, 475–485 (2002).
23. Lee, K. K. et al. Distinct functional domains in emerin bind lamin A and DNA-bridging protein BAF. *J. Cell Sci.* **114**, 4567–4573 (2001).
24. Wiskerchen, M. & Muesing, M. A. Human immunodeficiency virus type 1 integrase: Effects of mutations on viral ability to integrate, direct viral gene expression from unintegrated viral DNA templates, and sustain viral propagation in primary cells. *J. Virol.* **69**, 376–386 (1995).
25. Hazuda, D. J. et al. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287**, 646–650 (2000).
26. Emery, A. E. The muscular dystrophies. *Lancet* **359**, 687–695 (2002).
27. Jacque, J.-M., Triques, K. & Stevenson, M. Modulation of HIV-1 replication by RNAi. *Nature* **418**, 435–438 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Dauphin and J. Zhou for research support, B. Mellor for preparation of the figures, K. Departie and L. Smith for manuscript preparation, and A. Engelman for providing HIV-1 integrase mutants. The University of Massachusetts Center for AIDS Research and the AIDS Research Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health (NIH), provided assay support and reagents. This study was supported by grants from the NIH to M.S.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.S. (mario.stevenson@umassmed.edu).

LETTERS

A genome-wide *Drosophila* RNAi screen identifies DYRK-family kinases as regulators of NFAT

Yousang Gwack^{1,2*}, Sonia Sharma^{1,2*}, Julie Nardone^{1†}, Bogdan Tanasa¹, Alina Iuga^{1,2}, Sonal Srikanth^{1,2}, Heidi Okamura^{1,2†}, Diana Bolton^{1†}, Stefan Feske^{1,3}, Patrick G. Hogan¹ & Anjana Rao^{1,2}

Precise regulation of the NFAT (nuclear factor of activated T cells) family of transcription factors (NFAT1–4) is essential for vertebrate development and function¹. In resting cells, NFAT proteins are heavily phosphorylated and reside in the cytoplasm; in cells exposed to stimuli that raise intracellular free Ca^{2+} levels, they are dephosphorylated by the calmodulin-dependent phosphatase calcineurin and translocate to the nucleus¹. NFAT dephosphorylation by calcineurin is countered by distinct NFAT kinases, among them casein kinase 1 (CK1) and glycogen synthase kinase 3 (GSK3)^{1–5}. Here we have used a genome-wide RNA interference (RNAi) screen in *Drosophila*^{6,7} to identify additional regulators of the signalling pathway leading from Ca^{2+} –calcineurin to NFAT. This screen was successful because the pathways regulating NFAT subcellular localization (Ca^{2+} influx, Ca^{2+} –calmodulin–calcineurin signalling and NFAT kinases) are conserved across species^{8,9}, even though Ca^{2+} –regulated NFAT proteins are not themselves represented in invertebrates. Using the screen, we have identified DYRKs (dual-specificity tyrosine-phosphorylation regulated kinases) as novel regulators of NFAT. DYRK1A and DYRK2 counter calcineurin-mediated dephosphorylation of NFAT1 by directly phosphorylating the conserved serine-proline repeat 3 (SP-3) motif of the NFAT regulatory domain, thus priming further phosphorylation of the SP-2 and serine-rich region 1 (SRR-1) motifs by GSK3 and CK1, respectively. Thus, genetic screening in *Drosophila* can be successfully applied to cross evolutionary boundaries and identify new regulators of a transcription factor that is expressed only in vertebrates.

To validate the use of genome-wide RNAi screening in *Drosophila* to identify regulators of the Ca^{2+} –calcineurin–NFAT signalling pathway, we used an NFAT–GFP (green fluorescent protein) fusion protein¹⁰ containing the entire regulatory domain of NFAT1 (Fig. 1a). This domain bears >14 phosphorylated serines, 13 of which are dephosphorylated by calcineurin⁴. Five of the thirteen serines are located in the SRR-1 motif, which controls exposure of the nuclear localization sequence (NLS) and is a target for phosphorylation by CK1 (refs 4, 5); three are located in the SP-2 motif, which can be phosphorylated by GSK3 after a priming phosphorylation by protein kinase A (PKA)^{2,5}; and four are located in the SP-3 motif, for which a relevant kinase had yet to be identified at the time this study was initiated. The SP-2 and SP-3 motifs do not directly regulate the subcellular localization of NFAT1, but their dephosphorylation increases both the probability of NLS exposure and the affinity of NFAT for DNA^{1,4,5,11}. NFAT–GFP was correctly regulated in *Drosophila* S2R⁺ cells (Fig. 1b; see also Supplementary Fig. 1): it was phosphorylated and localized to the cytoplasm under resting conditions; it became dephosphorylated and translocated to the

nucleus with appropriate kinetics in response to Ca^{2+} store depletion with the sarcoplasmic/endoplasmic reticulum ATPase (SERCA) inhibitor thapsigargin; and its dephosphorylation and nuclear translocation were both sensitive to the calcineurin inhibitor cyclosporin

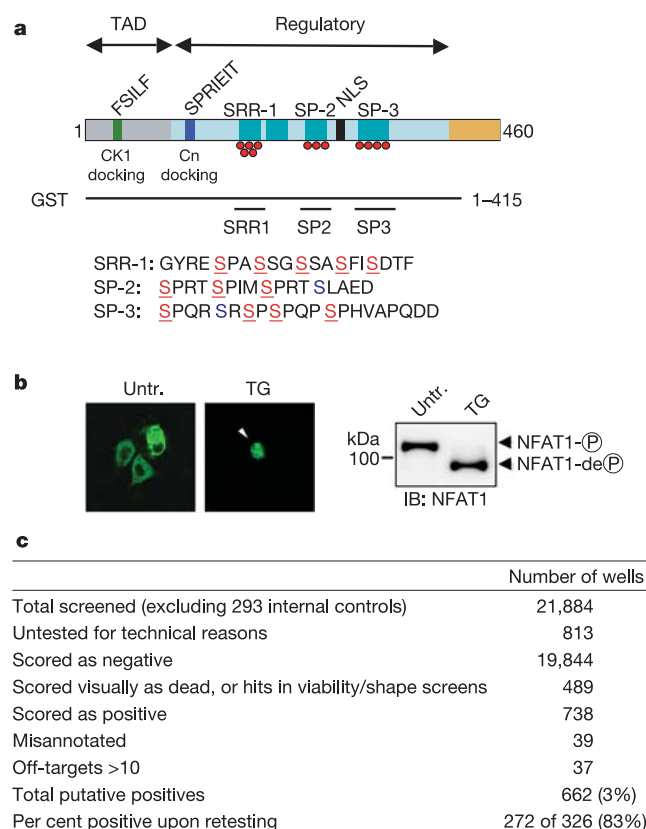


Figure 1 | The NFAT regulatory domain and the genome-wide RNAi screen in *Drosophila*. **a**, Diagram of the NFAT1 regulatory domain. Red circles or red and underlined residues show phosphorylated serine residues in the SRR-1, SP-2 and SP-3 motifs. NLS, nuclear localization signal; TAD, transactivation domain. **b**, NFAT is regulated by Ca^{2+} entry in *Drosophila* cells. S2R⁺ cells expressing GFP–NFAT1 were left untreated (Untr.) or treated with thapsigargin (TG, 1 μM , 30 min) and analysed by microscopy (left). Lysates were analysed by immunoblotting (IB) with anti-NFAT1 (right). P, phosphorylated; deP, dephosphorylated. **c**, Tabulation of the results of the primary screen.

¹The CBR Institute for Biomedical Research and the Departments of ²Pathology and ³Pediatrics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA. [†]Present addresses: Cell Signaling Technology, 166B Cummings Center, Beverly, Massachusetts 01915, USA (J.N.); AVEO Pharmaceuticals Inc., 75 Sidney Street, Cambridge, Massachusetts 02139, USA (H.O.); Department of Pediatrics, University of Washington, Seattle, Washington 98195, USA (D.B.).

*These authors contributed equally to this work.

A (CsA). S2R⁺ cells treated with limiting amounts of thapsigargin displayed intermediate phosphorylated forms of NFAT-GFP, most likely reflecting progressive dephosphorylation of serines within individual conserved motifs of the regulatory domain^{4,5} (Supplementary Fig. 1b). Depletion of the primary NFAT regulator, calcineurin, by RNAi in S2R⁺ cells inhibited thapsigargin-dependent dephosphorylation and nuclear import of NFAT-GFP (Supplementary Fig. 1c–e and Supplementary Table 2). Thus, the major pathways regulating NFAT phosphorylation and subcellular localization—store-operated Ca²⁺ influx, calcineurin activation and NFAT phosphorylation/dephosphorylation—are conserved in *Drosophila* and appropriately regulate vertebrate NFAT.

We performed a genome-wide RNAi screen⁷ on unstimulated S2R⁺ cells, and scored visually for aberrant nuclear localization of NFAT-GFP (Fig. 1c; see Methods and Supplementary Information). Positive candidates obtained in the screen (Supplementary Table 1) include Na⁺/Ca²⁺ exchangers and SERCA Ca²⁺ ATPases, the knockdown of which would be expected to increase basal levels of intracellular free Ca²⁺ ([Ca²⁺]_i)¹²; the scaffold protein Homer, which has been linked to Ca²⁺ influx and Ca²⁺ homeostasis¹³; stromal interaction molecule (STIM), a recently identified regulator of store-operated Ca²⁺ influx^{14–16} (Supplementary Fig. 2); and several protein kinases that control NFAT function either directly via phosphorylation or indirectly via basal [Ca²⁺]_i levels, calcineurin activity, or other kinases (Supplementary Tables 1 and 3). To identify kinases that directly phosphorylate the NFAT regulatory domain, we expressed Flag-tagged human homologues of selected *Drosophila* kinases in HEK293 cells, and tested anti-Flag immunoprecipitates in an *in vitro* kinase assay for their ability to phosphorylate a GST-NFAT1(1–415) fusion protein. Three kinases—protein kinase cGMP-dependent (PRKG1), DYRK2 and interleukin (IL)-1 receptor-associated kinase 4 (IRAK4)—showed strong activity in this assay (Fig. 2a, top; CK1-α and CK1-ε are positive controls). In cells, only DYRK2 countered the dephosphorylation of NFAT-GFP by calcineurin (Fig. 2b), even though both PRKG1 and DYRK2 were expressed at high levels (Fig. 2a, bottom). CD4⁺ T_H1 cells isolated from *Irak4*^{−/−} mice¹⁷ showed normal NFAT1 dephosphorylation, rephosphorylation and nuclear transport compared to control T_H1 cells (Supplementary Fig. 3a, b). We therefore focused on DYRK-family kinases as potential direct regulators of NFAT.

DYRKs constitute an evolutionarily conserved family of proline- or arginine-directed protein kinases belonging to the CMGC family of cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAPKs), GSK and CDK-like kinases (CLKs)^{18,19}. The DYRK family has multiple members (Supplementary Fig. 4a) that can be predominantly nuclear (DYRK1A and DYRK1B) or cytoplasmic (DYRK2–4 and homeodomain interacting protein kinase 3 (HIPK3)/DYRK6)²⁰. RT-PCR and western blotting suggested that DYRK1A and DYRK2 were major representatives of nuclear and cytoplasmic DYRKs, respectively, in Jurkat T cells (Supplementary Fig. 4b)—a conclusion supported by several additional observations. First, overexpression of DYRK2 prevented dephosphorylation of NFAT-GFP after ionomycin treatment (Fig. 2b, lanes 5–8); overexpression of wild-type DYRK2, but not a kinase-dead mutant of DYRK2 (refs 21, 22), prevented NFAT nuclear localization in thapsigargin-treated cells (Fig. 2c). The slower-migrating form of NFAT in lanes 7 and 8 of Fig. 2b might indicate that DYRK2, a proline-directed kinase¹⁸, acted in part by phosphorylating the SPRIETP docking sequence on NFAT1 (refs 1, 10) and thereby blocking the calcineurin–NFAT interaction. However, DYRK was still effective when the potential DYRK phosphorylation sites in the docking sequence were eliminated by substituting HPVIVITGP for SPRIETPS¹⁰ (Supplementary Fig. 5a). Second, both wild-type and kinase-dead DYRK co-immunoprecipitated with NFAT1 (Supplementary Fig. 5b). Third, depletion of the DYRK-family candidate CG40478 in S2R⁺ cells did not affect—and DYRK2 overexpression in Jurkat T cells only slightly diminished—Ca²⁺

mobilization in response to thapsigargin (Supplementary Fig. 6a, b). Finally, and most importantly, depletion of endogenous DYRK1A with DYRK1A-specific short interfering RNAs (siRNA) in HeLa cells stably expressing NFAT-GFP increased the rate and extent of NFAT1 dephosphorylation and nuclear import while slowing rephosphorylation and nuclear export (Fig. 3a, b). These results show that DYRK1A and DYRK2 are physiological negative regulators of NFAT activation in cells. The absence of basal NFAT dephosphorylation in DYRK1A-depleted cells may reflect both the expression of other DYRK family members in human cells (Supplementary Fig. 4b) and the predominantly nuclear localization of DYRK1A (refs 19, 20).

DYRKs are direct NFAT1 kinases that selectively phosphorylate the SP-3 motif, but nevertheless control the overall phosphorylation

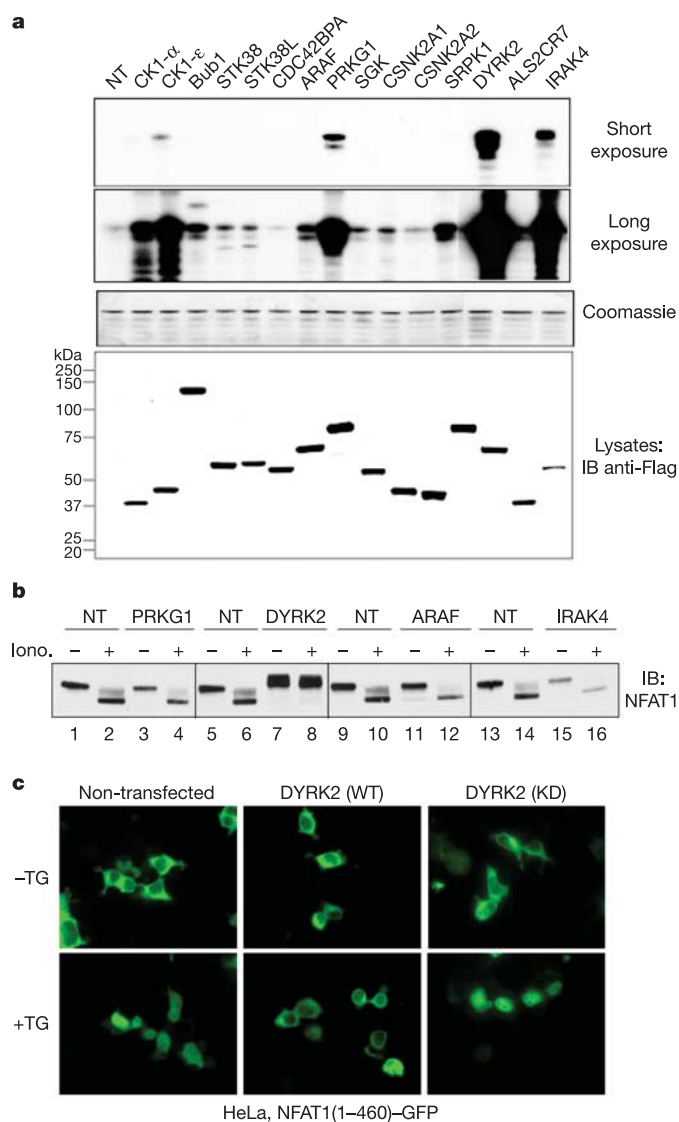


Figure 2 | Screening for NFAT phosphorylation: DYRK kinase negatively regulates NFAT activation. **a**, Phosphorylation of the NFAT regulatory domain. Flag immunoprecipitates of human homologues of selected *Drosophila* kinases were tested for GST-NFAT1(1–415) phosphorylation, as assessed by autoradiography (top and middle panels). Kinase expression was verified by immunoblotting using anti-Flag antibody (bottom panel). NT, not transfected. **b**, DYRK2 overexpression blocks NFAT-GFP dephosphorylation in ionomycin-treated cells, assessed by immunoblotting of cell lysates with anti-NFAT1. **c**, Overexpression of wild-type (WT), but not kinase-dead (KD), DYRK2 blocks NFAT-GFP nuclear translocation. NFAT-GFP localization (green) was examined by fluorescence microscopy in untreated (–TG) and thapsigargin-treated (+TG) cells.

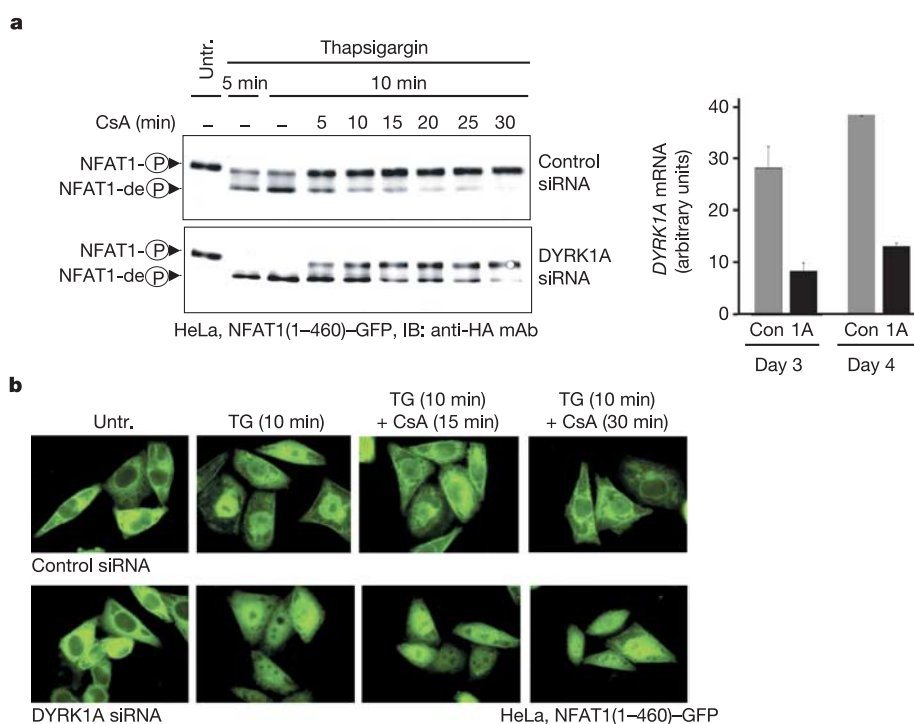


Figure 3 | Depletion of endogenous DYRK1A potentiates NFAT activation. **a**, Depletion of DYRK1A delays re-phosphorylation of nuclear NFAT. HeLa cells stably expressing NFAT-GFP were transfected with control (Con) or DYRK1A-specific (1A) siRNAs, and treated with 1 μ M thapsigargin (TG) followed by 20 nM CsA. Phosphorylation status of NFAT-GFP was assessed by western blotting with anti-HA (left). DYRK1A messenger RNA levels were assessed 3 and 4 days after transfection by real-time PCR (right). Results show average $\pm$ s.d. of three independent experiments. **b**, Depletion of DYRK1A delays NFAT nuclear export. NFAT-GFP localization (green) was examined by fluorescence microscopy in siRNA-transfected cells.

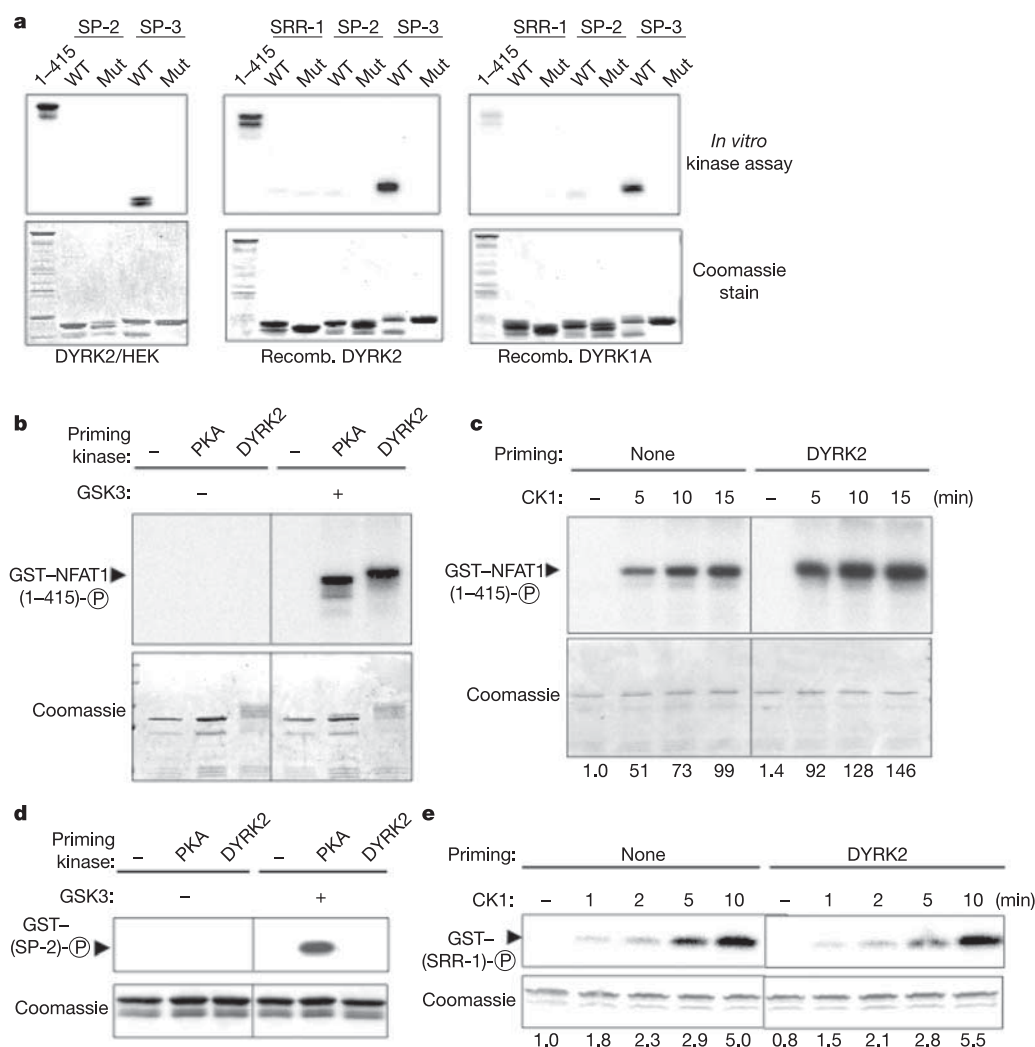


Figure 4 | Pre-phosphorylation by DYRK primes the NFAT regulatory domain for subsequent phosphorylation by GSK3 and CK1. **a**, Immunopurified DYRK2, recombinant DYRK2 or recombinant DYRK1A was incubated with GST-NFAT1(1-415); with the indicated GST-peptide fusion GST-SRR-1, GST-SP-2 or GST-SP-3 (WT); or with a corresponding GST-peptide fusion (Mut) carrying serine-to-alanine replacements at the residues known to be phosphorylated in cells (Fig. 1a). **b**, DYRK2 and PKA prime for GSK3 phosphorylation of the NFAT1 regulatory domain (quantification below). **c**, DYRK2 potentiates CK1 phosphorylation of the NFAT1 regulatory domain (quantification below). **d**, DYRK2 does not prime for GSK3 phosphorylation of the SP-2 motif. **e**, DYRK2 does not prime for CK1 phosphorylation of the SRR-1 motif (quantification below).

of NFAT1 (Fig. 4). Flag-tagged DYRK2 expressed in HEK cells, as well as bacterially expressed recombinant DYRK1A and DYRK2, phosphorylated peptides corresponding to the SP-3 motif of the NFAT1 regulatory domain *in vitro*, but did not phosphorylate SRR-1 or SP-2 peptides or an SP-3 peptide with serine to alanine substitutions in the known phosphoserine residues⁴ (Fig. 4a). Two serine residues (underlined) in the SP-3 motif (SPQR SRSPSPQSPHVPQDD) fit the known sequence preference of DYRK kinases (R(x)xx(S/T)(P/V))^{23–25}, and both are known to be phosphorylated in cells⁴ (Fig. 1a). DYRK is reported to prime for GSK3-mediated phosphorylation of eukaryotic initiation factor 2B-ε

(eIF2B-ε) and the microtubule-associated protein tau²⁵, as well as for GSK3- and CK1-mediated phosphorylation of OMA-1 (refs 26 and 27). We therefore investigated whether DYRK kinases could also prime for GSK3- and CK1-mediated phosphorylation of NFAT1. Pre-phosphorylation of the NFAT1 regulatory domain by DYRK2 led to robust phosphorylation by GSK3 and induced the mobility shift characteristic of phosphorylation of the SP-2 and SP-3 motifs⁴; this shift was not observed after pre-phosphorylation by PKA (Fig. 4b). Furthermore, pre-phosphorylation by DYRK2 accelerated CK1-mediated phosphorylation of GST–NFAT1(1–415) by at least two-fold (Fig. 4c). In contrast, DYRK2 did not prime phosphorylation of the SP-2 peptide by GSK3 (Fig. 4d) nor the SRR-1 peptide by CK1 (Fig. 4e), consistent with the fact that neither motif is a substrate for DYRK (Fig. 4a). This ‘discontiguous’ priming mechanism is distinct from conventional priming, which requires phosphorylation at +4 and –3 for GSK3 and CK1, respectively^{28,29}. A less likely interpretation is that the conventional priming sites for CK1 and GSK3 are efficiently phosphorylated by DYRK in the context of the GST–NFAT1(1–415) protein, although they are not phosphorylated in the peptide context.

We used the kinase-dead mutant of DYRK2 to show that DYRK regulates the transcriptional activity of NFAT. Wild-type DYRK2 strongly diminished NFAT-dependent activity, whereas the kinase-dead mutant increased NFAT-dependent luciferase activity of the *IL2* promoter (Fig. 5a); an NFAT-activating protein 1 (AP-1) reporter (data not shown); or an AP-1-independent promoter, the κ3 site of the tumour-necrosis factor-α (TNF-α) promoter (Supplementary Fig. 6c). Similarly, wild-type DYRK2 diminished production of endogenous IL-2 by stimulated Jurkat T cells, whereas kinase-dead DYRK2 had the opposite effect (Fig. 5b, c).

Our data indicate that DYRK is a key kinase that regulates NFAT1 phosphorylation. DYRK, GSK3 and CK1 target completely distinct motifs of the NFAT1 regulatory domain, but DYRK-mediated phosphorylation of the SP-3 motif primes for further phosphorylation of the distinct SRR-1 and SP-2 motifs by CK1 and GSK3, respectively, thus facilitating complete phosphorylation and deactivation of NFAT1. This mechanism, which we term ‘discontiguous priming’, is reminiscent of that recently proposed for *Caenorhabditis elegans* oocyte maturation protein 1 (OMA-1), in which phosphorylation of Thr 239 by the DYRK-family kinase minibrain kinase 2 (MBK-2) potentiates GSK3-mediated phosphorylation of Thr 339 (refs 26, 27). It is likely that DYRK2, DYRK3 and DYRK4, which are localized to the cytoplasm¹⁹, function primarily as ‘maintenance’ kinases that sustain the phosphorylation status of cytoplasmic NFAT in resting cells, whereas DYRK1A and DYRK1B, which are localized to the nucleus^{19,20}, re-phosphorylate nuclear NFAT and promote its nuclear export. Notably, NFAT dephosphorylation may also proceed through a sequential mechanism, with dephosphorylation of the SRR-1 motif promoting dephosphorylation of the SP-2 and SP-3 motifs by increasing their accessibility to calcineurin⁴. DYRK1A and the endogenous calcineurin regulator *DSCR1/RCN/calciressin-1* are both localized to the Down’s syndrome critical region on human chromosome 21; thus, overexpression of these negative regulators of NFAT could contribute—by inhibiting NFAT activation—to the neurological and immunological developmental anomalies observed in individuals with chromosome 21 trisomy³⁰.

We have shown that genome-wide RNAi screening in *Drosophila* is a valid and powerful strategy for exploring novel aspects of signal transduction in mammalian cells, provided that key members of the signalling pathway are evolutionarily conserved and represented in the *Drosophila* genome. We have used the method to identify conserved regulators of the purely vertebrate transcription factor NFAT; to our knowledge, this is the first example of a genome-wide RNAi screen that crosses evolutionary boundaries in this manner. It is likely that conserved aspects of the regulation of other mammalian processes will also be successfully defined by developing assays in *Drosophila* cells.

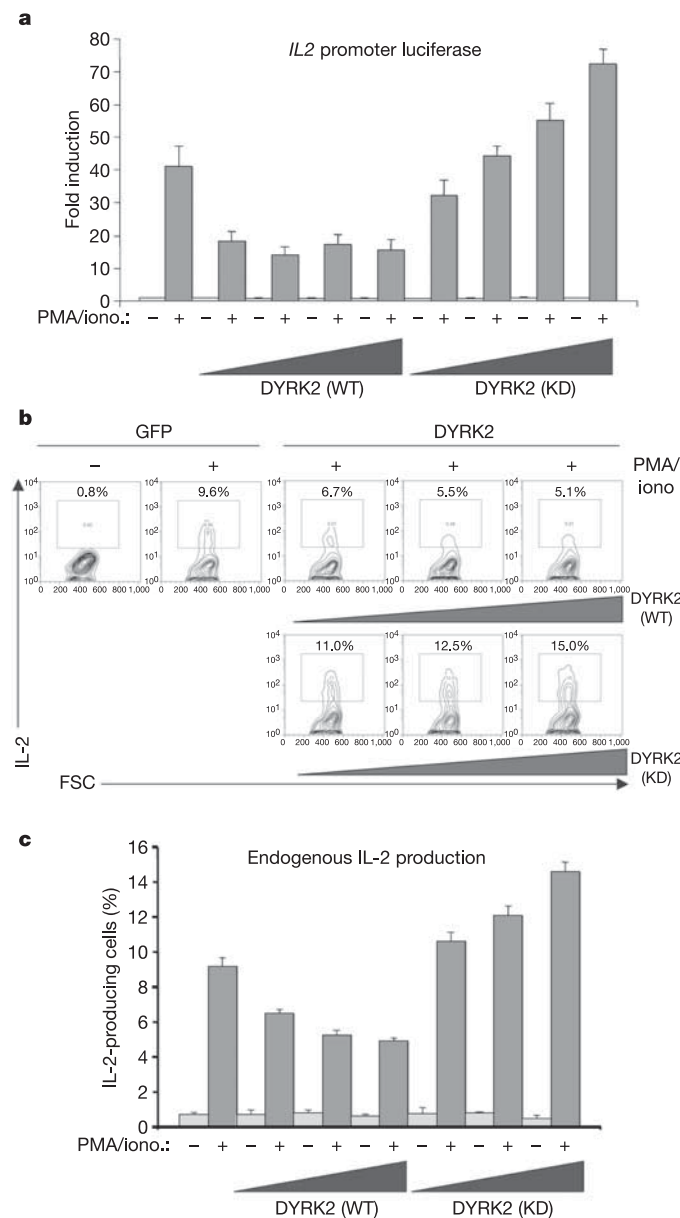


Figure 5 | DYRK2 inhibits IL-2 transcription. **a**, DYRK2 inhibits *IL2* promoter activity. Jurkat T cells, co-transfected with *Renilla* and *IL2* promoter-driven luciferase reporters and wild-type or kinase-dead DYRK2 expression plasmids (5, 10, 15, 20 μg), were left untreated or stimulated with PMA/ionomycin (iono). Results show average ± s.d. of three independent experiments. **b**, DYRK2 inhibits endogenous IL-2 expression. IL-2 expression was evaluated by intracellular cytokine staining and flow cytometry in GFP⁺ Jurkat T cells transfected with GFP and wild-type or kinase-dead DYRK2 expression plasmids (10, 20, 30 μg). **c**, Quantification of **b** (average ± s.d. of three independent experiments).

METHODS

Detailed methods are presented in Supplementary Information.

The genome-wide primary screen. Methods were adapted from refs 6 and 7. A total of 10^4 S2R⁺ cells were added to each well containing 0.25 µg double-stranded (ds)RNAs in 10 µl serum-free medium and incubated for 1 h at 26 °C. The cells were then transiently transfected with NFAT1(1–460)–GFP expression plasmid¹⁰ (10 ng) in Schneider's medium (Invitrogen) (30 µl). After incubation for 48–72 h at 26 °C, the cells were fixed and stained with 4,6-diamidino-2-phenylindole (DAPI), and the coincident GFP and DAPI images were acquired by an automated camera from three different locations in each well. A total of 58 384-well plates were analysed, containing a total of 21,884 wells into which individual dsRNAs had been arrayed. The confirmatory screening on the 699 potentially positive candidates from the primary screen was performed essentially as described for the primary screen, except that S2R⁺ cells stably transfected with NFAT1(1–460)–GFP were used.

Received 14 October 2005; accepted 1 February 2006.

Published online 1 March 2006.

- Hogan, P. G., Chen, L., Nardone, J. & Rao, A. Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes Dev.* **17**, 2205–2232 (2003).
- Beals, C. R., Sheridan, C. M., Turck, C. W., Gardner, P. & Crabtree, G. R. Nuclear export of NF-ATc enhanced by glycogen synthase kinase-3. *Science* **275**, 1930–1934 (1997).
- Zhu, J. *et al.* Intramolecular masking of nuclear import signal on NF-AT4 by casein kinase I and MEK1. *Cell* **93**, 851–861 (1998).
- Okamura, H. *et al.* Concerted dephosphorylation of the transcription factor NFAT1 induces a conformational switch that regulates transcriptional activity. *Mol. Cell* **6**, 539–550 (2000).
- Okamura, H. *et al.* A conserved docking motif for CK1 binding controls the nuclear localization of NFAT1. *Mol. Cell Biol.* **24**, 4184–4195 (2004).
- Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).
- Armknrecht, S. *et al.* High-throughput RNA interference screens in *Drosophila* tissue culture cells. *Methods Enzymol.* **392**, 55–73 (2005).
- Myers, E. W. *et al.* A whole-genome assembly of *Drosophila*. *Science* **287**, 2196–2204 (2000).
- Yeromin, A. V., Roos, J., Stauderman, K. A. & Cahalan, M. D. A store-operated calcium channel in *Drosophila* S2 cells. *J. Gen. Physiol.* **123**, 167–182 (2004).
- Aramburu, J. *et al.* Affinity-driven peptide selection of an NFAT inhibitor more selective than cyclosporin A. *Science* **285**, 2129–2133 (1999).
- Neal, J. W. & Clipstone, N. A. Glycogen synthase kinase-3 inhibits the DNA binding activity of NFATc. *J. Biol. Chem.* **276**, 3666–3673 (2001).
- Carafoli, E. The calcium signalling saga: tap water and protein crystals. *Nature Rev. Mol. Cell Biol.* **4**, 326–332 (2003).
- Roderick, H. L. & Bootman, M. D. Calcium influx: is Homer the missing link? *Curr. Biol.* **13**, R976–R978 (2003).
- Roos, J. *et al.* STIM1, an essential and conserved component of store-operated Ca²⁺ channel function. *J. Cell Biol.* **169**, 435–445 (2005).
- Liou, J. *et al.* STIM is a Ca²⁺ sensor essential for Ca²⁺-store-depletion-triggered Ca²⁺ influx. *Curr. Biol.* **15**, 1235–1241 (2005).
- Zhang, S. L. *et al.* STIM1 is a Ca²⁺ sensor that activates CRAC channels and migrates from the Ca²⁺ store to the plasma membrane. *Nature* **437**, 902–905 (2005).
- Suzuki, N. *et al.* Severe impairment of IL-1 and TLR signalling in mice lacking IRAK-4. *Nature* **416**, 750–756 (2002).
- Kannan, N. & Neuwald, A. F. Evolutionary constraints associated with functional specificity of the CMGC protein kinases MAPK, CDK, GSK, SRPK, DYRK, and CK2α. *Protein Sci.* **13**, 2059–2077 (2004).
- Lochhead, P. A., Sibbet, G., Morrice, N. & Cleghon, V. Activation-loop autophosphorylation is mediated by a novel transitional intermediate form of DYRKs. *Cell* **121**, 925–936 (2005).
- Becker, W. & Joost, H. G. Structural and functional characteristics of Dyrk, a novel subfamily of protein kinases with dual specificity. *Prog. Nucleic Acid Res. Mol. Biol.* **62**, 1–17 (1999).
- Kentrup, H. *et al.* Dyrk, a dual specificity protein kinase with unique structural features whose activity is dependent on tyrosine residues between subdomains VII and VIII. *J. Biol. Chem.* **271**, 3488–3495 (1996).
- Wiechmann, S. *et al.* Unusual function of the activation loop in the protein kinase DYRK1A. *Biochem. Biophys. Res. Commun.* **302**, 403–408 (2003).
- Himpel, S. *et al.* Specificity determinants of substrate recognition by the protein kinase DYRK1A. *J. Biol. Chem.* **275**, 2431–2438 (2000).
- Campbell, L. E. & Proud, C. G. Differing substrate specificities of members of the DYRK family of arginine-directed protein kinases. *FEBS Lett.* **510**, 31–36 (2002).
- Woods, Y. L. *et al.* The kinase DYRK phosphorylates protein-synthesis initiation factor eIF2Bε at Ser539 and the microtubule-associated protein tau at Thr212: potential role for DYRK as a glycogen synthase kinase 3-priming kinase. *Biochem. J.* **355**, 609–615 (2001).
- Nishi, Y. & Lin, R. DYRK2 and GSK-3 phosphorylate and promote the timely degradation of OMA-1, a key regulator of the oocyte-to-embryo transition in *C. elegans*. *Dev. Biol.* **288**, 139–149 (2005).
- Shirayama, M. *et al.* The conserved kinases CDK-1, GSK-3, KIN-19, and MBK-2 promote OMA-1 destruction to regulate the oocyte-to-embryo transition in *C. elegans*. *Curr. Biol.* **16**, 47–55 (2006).
- Harwood, A. J. Regulation of GSK-3: a cellular multiprocessor. *Cell* **105**, 821–824 (2001).
- Flotow, H. *et al.* Phosphate groups as substrate determinants for casein kinase I action. *J. Biol. Chem.* **265**, 14264–14269 (1990).
- Bhattacharyya, A. & Svendsen, C. N. Human neural stem cells: a new tool for studying cortical development in Down's syndrome. *Genes Brain Behav.* **2**, 179–186 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Mathey-Prevot, N. Perrimon, S. Armknrecht, N. Ramadan, J. Murphy, K. Richardson, M. Booker and staff at the *Drosophila* RNAi Screening Centre at Harvard Medical School for valuable assistance with the screens. We also thank M. Dziadek for the gift of antibody against STIM1, and W. C. Yeh and T. Mak for sending spleen and lymph nodes isolated from *Irak4*^{+/−} and *Irak4*^{−/−} mice. This work was supported by NIH grants to A.R. and S.F. S. Sharma is supported by a postdoctoral fellowship from the Canadian Institutes of Health Research.

Author Contributions The *Drosophila* RNAi screen was designed, validated and performed by Y.G., and positive candidates were identified through visual screening performed by Y.G., J.N. and H.O. S. Sharma identified DYRK as a functional NFAT kinase through biochemical and functional analyses of the human homologues of candidate *Drosophila* kinases, drawing in part on unpublished data and reagents provided by H.O. Selected other candidates were investigated in detail by Y.G., S. Sharma, D.B. and S. Srikanth. A.I. and S.F. performed measurements of intracellular calcium concentration in mammalian cells. J.N. and B.T. were responsible for bioinformatic analysis. P.G.H. and A.R. provided overall direction and supervised project planning and execution.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.R. (arao@cbr.med.harvard.edu)

Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit

Thorsten Schäfer¹, Bohumil Maco², Elisabeth Petfalski³, David Tollervey³, Bettina Böttcher⁴, Ueli Aebi² & Ed Hurt¹

The formation of eukaryotic ribosomes is a multistep process that takes place successively in the nucleolar, nucleoplasmic and cytoplasmic compartments^{1–4}. Along this pathway, multiple pre-ribosomal particles are generated, which transiently associate with numerous non-ribosomal factors before mature 60S and 40S subunits are formed^{5–12}. However, most mechanistic details of ribosome biogenesis are still unknown. Here we identify a maturation step of the yeast pre-40S subunit that is regulated by the protein kinase Hrr25 and involves ribosomal protein Rps3. A high salt concentration releases Rps3 from isolated pre-40S particles but not from mature 40S subunits. Electron microscopy indicates that pre-40S particles lack a structural landmark present in mature 40S subunits, the ‘beak’. The beak is formed by the protrusion of 18S ribosomal RNA helix 33, which is in close vicinity to Rps3. Two protein kinases Hrr25 and Rio2 are associated with pre-40S particles. Hrr25 phosphorylates Rps3 and the 40S synthesis factor Enp1. Phosphorylated Rps3 and Enp1 readily dissociate from the pre-ribosome, whereas subsequent dephosphorylation induces formation of the beak structure and salt-resistant integration of Rps3 into the 40S subunit. *In vivo* depletion of Hrr25 inhibits growth and leads to the accumulation of immature 40S subunits that contain unstably bound Rps3.

We conclude that the kinase activity of Hrr25 regulates the maturation of 40S ribosomal subunits.

To analyse the maturation of pre-40S ribosomal subunits, we assessed their interactions with non-ribosomal factors. Pre-40S particles were isolated by tandem affinity purification (TAP) of the associated non-ribosomal bait protein Rio2-TAP¹² and subjected to gel filtration in the presence of a high salt concentration (100 mM MgCl₂). Under these conditions all associated non-ribosomal factors (namely Rio2, Tsr1, Ltv1, Enp1, Nob1, Hrr25, Dim1 and Dim2) were dissociated from the 40S subunit (Fig. 1a). The ribosomal protein Rps3 was also released, but the bulk of small subunit proteins including Rps8 were not dissociated (Fig. 1a). In contrast, mature 40S subunits were not disrupted by 100 mM MgCl₂; in particular, Rps3 remained associated (Supplementary Fig. S1). Gel-filtration chromatography of the salt-extracted pre-40S particle further revealed that Enp1, Ltv1 and Rps3 were eluted together in intermediate fractions, which is indicative of complex formation (Fig. 1a). The remaining 40S synthesis factors (for example Tsr1 and Rio2) were eluted in later fractions, indicating that they became monomeric.

Ltv1 was previously shown to be associated with pre-40 subunits¹³, whereas Enp1 is present in both early 90S and late 40S pre-ribosomes¹². Consistent with these findings was our observation

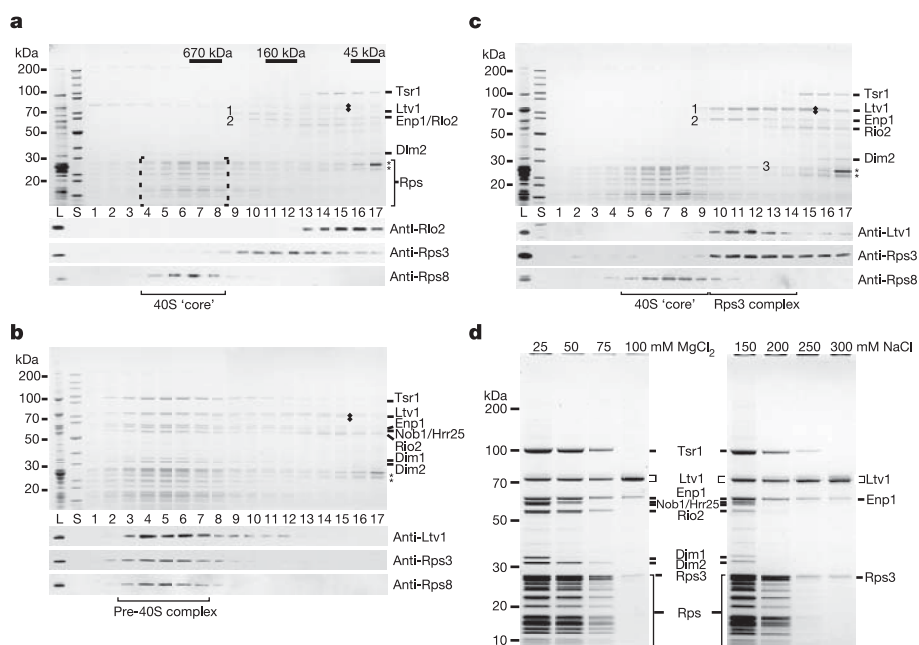


Figure 1 | Ltv1, Enp1 and Rps3 form an extractable pre-ribosomal subcomplex.

a–c, Gel-filtration profiles of MgCl₂-treated pre-40S particles. Rio2-TAP (**a**) or Ltv1-TAP (**b, c**) purifications were separated under conditions of high (100 mM MgCl₂; **a, c**) or low (10 mM MgCl₂; **b**) salt concentration. Input (L), standard (S) and column fractions (1–17) were analysed by SDS-PAGE and Coomassie staining, or by western blotting with anti-CBP (bait), anti-Rps3 and anti-Rps8 antibodies. Band 1, Ltv1; band 2, Enp1; band 3, Rps3; asterisks, TEV (Tobacco Etch Viral protease); diamonds, Hsp70s. A 40S ‘core’ particle is eluted at about 10⁶ Da. **d**, Affinity purifications of Ltv1-TAP from lysates at the indicated MgCl₂ or NaCl concentrations. Eluates were analysed by SDS-PAGE and Coomassie staining.

¹Biochemie-Zentrum der Universität Heidelberg, Im Neuenheimer Feld 328, 69120 Heidelberg, Germany. ²M. E. Müller Institute for Structural Biology, Biozentrum, Universität Basel, CH-4056 Basel, Switzerland. ³Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh EH9 3JR, UK. ⁴EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

that Ltv1-TAP co-precipitated a late pre-40S particle indistinguishable from the Rio2-TAP particle (Fig. 1b). When the Ltv1 particle was incubated with 100 mM MgCl₂, Rps3 and all non-ribosomal factors were dissociated, and Ltv1, Enp1 and Rps3 were again eluted together (Fig. 1c).

To show more directly that Ltv1 and Enp1 form a salt-resistant complex with Rps3, Ltv1-TAP was affinity-purified under different salt conditions. After purification in 100 mM MgCl₂ or 300 mM NaCl (Fig. 1d), most non-ribosomal factors and ribosomal proteins were dissociated from Ltv1-TAP, whereas Enp1 and Rps3 remained bound. Taken together, these data showed that Ltv1, Enp1 and Rps3 exist in a salt-resistant complex that can be released from pre-40S subunits, whereas Rps3 cannot be extracted from mature 40S subunits under these conditions.

The observation that Rps3 showed differential association with pre-40S and mature 40S subunits prompted us to compare their structures. The eukaryotic small subunit consists of three characteristic subdomains referred to as 'head', 'body' and 'platform'. Protruding rRNA turns and helices provide additional characteristic features in the 40S subunit structure, termed the 'beak', 'shoulder' and 'feet'^{14–16} (see also Fig. 2c). Comparison of the X-ray structure of the prokaryotic 30S subunit¹⁷ with cryoelectron-microscopy-based reconstruction of the yeast ribosome has allowed the structure of the mature 40S subunit to be modelled at atomic resolution¹⁶.

We initially compared negatively stained pre-40S subunits with mature 40S subunits by electron microscopy (Fig. 2a, and Supplementary Fig. S2a). Both particles exhibit the typical 'head', 'platform' and 'body' domains. However, the pre-40S ribosomal particle lacks the prominent 'beak', which is a protrusion of helix 33 of the 18S rRNA (Fig. 2a, c). Moreover, additional mass is visible adjacent to the 'platform' of the pre-40S particle, which could represent non-ribosomal factors (see also Supplementary Fig. S2a).

Next we performed cryoelectron microscopy to further resolve the structural differences between nascent and mature 40S subunits (Fig. 2b, and Supplementary Fig. S2b). Three-dimensional reconstruction of the pre-40S particle showed the major structural landmarks of the small subunit, but the 'beak' was not visible (Fig. 2b, c). However, an elongated structure was revealed in the pre-40S subunit, which emerged from the head region and wrapped around in a clockwise orientation. It is possible that this structure is the progenitor of the genuine 'beak' that protrudes from the mature 40S subunit. We also observed an additional structure on the opposite side of the head, which could be a flipped beak (see also Supplementary Fig. S2b and animated rotations). Notably, the prokaryotic homologue of Rps3 (called S3) is located at the base of a structure within the 30S subunit, which corresponds to the eukaryotic beak¹⁷. Three-dimensional reconstruction of the eukaryotic 40S subunit places yeast Rps3 in a similar position¹⁶, indicating that it might define the exposed position of eukaryotic helix 33 within the mature 40S subunit (see also Fig. 2c).

The evidence that Rps3 is not incorporated into the pre-40S subunit in its final conformation would be consistent with a role for Rps3 in 40S biogenesis. Depletion of the essential Rps3 from yeast cells resulted in a strongly reduced export rate of 20S rRNA¹⁸ and nuclear accumulation of pre-40S subunits (Supplementary Fig. S3). Moreover, 20S and 35S pre-rRNA were increased and 27SA₂ pre-rRNA was markedly decreased in the *GAL1::RPS3* depletion mutant (Supplementary Fig. S3).

We next addressed how the salt-resistant incorporation of Rps3 into the 40S subunit and formation of the 'beak' structure might be regulated. Two protein kinases, Rio2 and Hrr25, are associated with the purified pre-40S subunits¹², indicating that phosphorylation might have a function in small subunit maturation. Notably, when Ltv1-TAP was isolated from yeast lysates in the presence of phosphatase inhibitors, two bands, Ltv1 and Enp1, were shifted and

migrated slower in the SDS-polyacrylamide gel when compared with a preparation performed in the absence of phosphatase inhibitors (Supplementary Fig. S4a). These data indicate that Ltv1 and Enp1 are phosphorylated *in vivo*.

To assess whether these proteins can undergo phosphorylation *in vitro*, pre-40S particles were purified without phosphatase inhibitors to allow the removal of phosphate by endogenous phosphatases. Incubation with ATP at 4 °C induced a shift in the mobility of the Ltv1 and Enp1 bands on the SDS-polyacrylamide gel, indicating that they were phosphorylated *in vitro* (Fig. 3a). When the pre-40S subunits were incubated with ATP at physiological temperature (30 °C), Ltv1, Enp1 and Rps3 were phosphorylated and were also dissociated from the pre-ribosomes (Fig. 3b). In contrast, other non-ribosomal factors were not dissociated. The bands corresponding to phosphorylated Ltv1, Enp1 and Rps3 were eluted together in intermediate fractions of the gel-filtration column, indicating that they remained as a complex (Fig. 3b, and Supplementary Fig. S4b). In contrast, incubation of the pre-40S particle at 30 °C in the absence of ATP did not cause dissociation of Ltv1, Enp1 or Rps3 from the pre-40S particle (Fig. 3c, and Supplementary Fig. S4b). These data

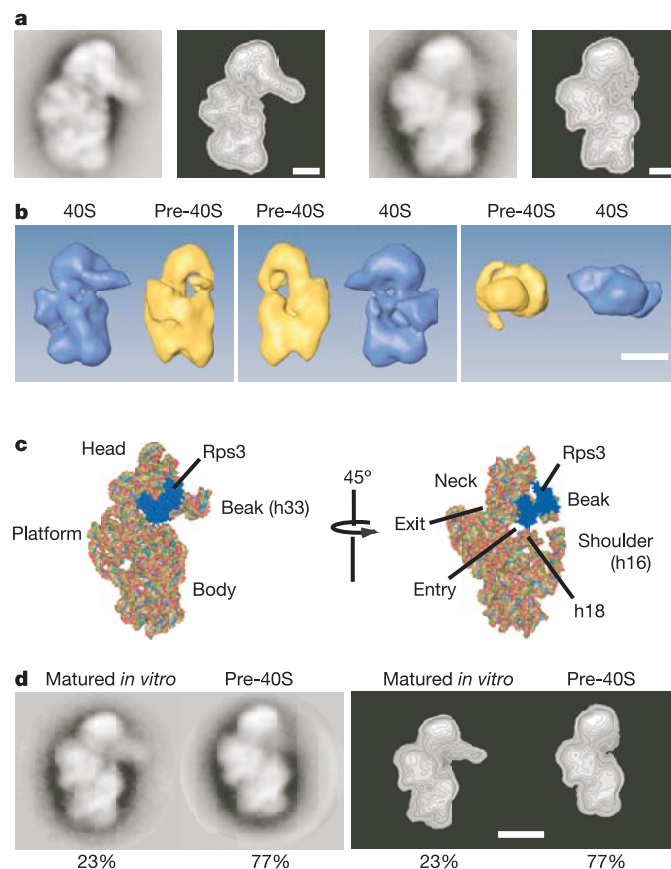


Figure 2 | Structural comparison of pre-40S and mature 40S subunits.

a, Projection averages (left) and electron-density contour maps (right) of negatively stained 40S (left pair) and pre-40S (right pair) subunits. Scale bars, 5 nm. **b**, Three-dimensional reconstruction of mature and pre-40S particles from cryoelectron micrographs. Depicted are orientations from the solvent (left), intersubunit (centre) and top (right) sides. Scale bar, 10 nm. **c**, Tertiary structure of yeast 18S rRNA with superimposed Rps3 (blue) as described in www.rcsb.org (Protein Data Bank, 1S1H). Major structural landmarks of the 40S subunit are depicted. **d**, Development of the beak structure within isolated pre-40S particles on phosphorylation and subsequent dephosphorylation. In 23% of the isolated pre-40S particles (affinity-purified by means of Rio2-TAP) beak formation was observed by electron microscopy after negative staining. Left, class averages; right, electron-density contour maps. Scale bar, 10 nm.

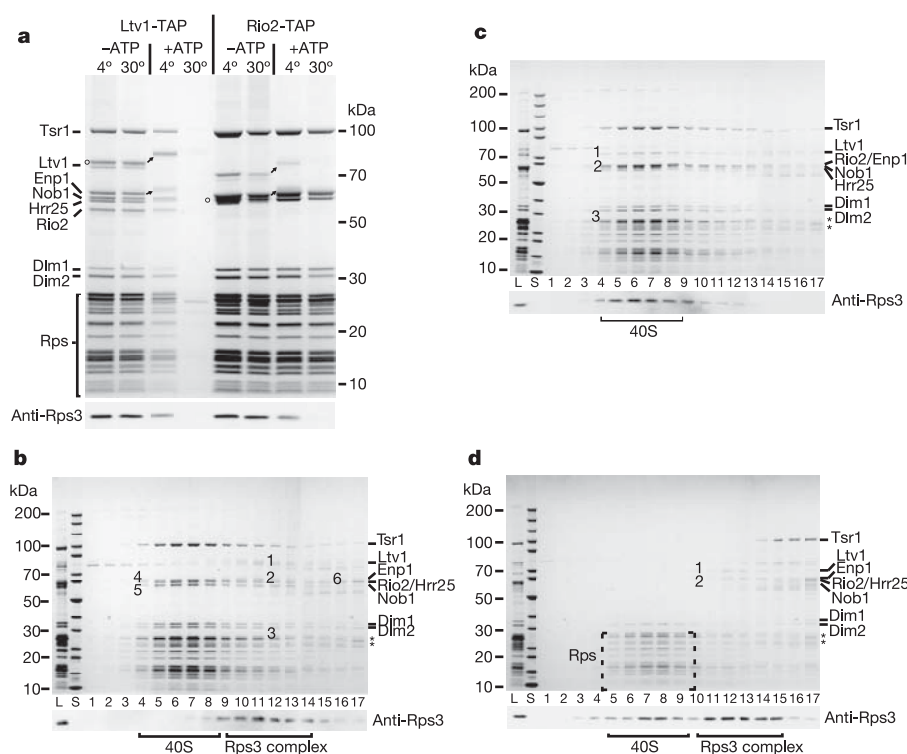


Figure 3 | Phosphorylation state induces 40S subunit maturation. **a**, ATP-induced phosphorylation (4 °C) and dissociation of Ltv1 and Enp1 (30 °C) from isolated pre-40S subunits. The indicated eluates of Ltv1-TAP and Rio2-TAP preparations were analysed by SDS-PAGE and Coomassie staining or western blotting (anti-Rps3 antibody). Circles indicate bait proteins; arrows identify ATP-induced band shifts. Ltv1-TAP could not be eluted from IgG-beads on treatment with ATP at 30 °C. **b-d**, ATP-dependent dissociation of Ltv1-Enp1-Rps3 complex from pre-40S particles. Incubation of Rio2-TAP preparations with 1 mM ATP (**b**) or no ATP (**c**) for 30 min at 30 °C and analysis by gel filtration. **d**, A phosphorylation-dephosphorylation cycle induces salt-resistant association of Rps3 with the pre-40S subunit. Gel-filtration profiles of Rio2-TAP preparations treated with 10 mM (**b**, **c**) or 100 mM (**d**) MgCl₂, analysed by SDS-PAGE and Coomassie staining or western blotting (anti-Rps3 antibody). Bands are identified as follows in **b-d**: 1, Ltv1; 2, Enp1; 3, Rps3; 4, Rio2; 5, Nob1; 6, Hrr25; asterisks, TEV protease.

indicate that treatment with ATP at physiological temperature induced the phosphorylation and subsequent dissociation of Ltv1, Enp1 and Rps3 from the 40S pre-ribosome.

We next tested whether phosphorylation of the pre-40S particle was followed by a dephosphorylation step that triggers further pre-40S subunit maturation. No specific phosphatases that participate in ribosome biogenesis have been identified, and ATP-treated pre-40S particles were therefore incubated with λ -phosphatase. Subunit maturation was monitored by gel-filtration chromatography and electron microscopy. Biochemical analysis showed that a significant fraction of Rps3 (about 25%) became associated with the mature 40S subunits in a salt-resistant manner after a round of phosphorylation

followed by dephosphorylation (Fig. 3d). Consistent with this was our EM examination, which revealed that 23% of negatively stained pre-40S particles showed the characteristic beak structure, after phosphorylation-dephosphorylation treatment (Fig. 2d). In contrast, Rps3 was not stably associated with 40S subunits if either the phosphorylation or dephosphorylation step was applied alone (Supplementary Fig. S5a, b).

To identify the kinase(s) responsible for phosphorylation of the pre-40S components, the two candidate kinases, RIO2 and HRR25, were repressed in yeast with the use of the regulated *GAL1* promoter (Supplementary Figs S6 and S7). Rio2 has been reported to be involved in the processing of 20S pre-rRNA¹⁹ and nuclear 40S

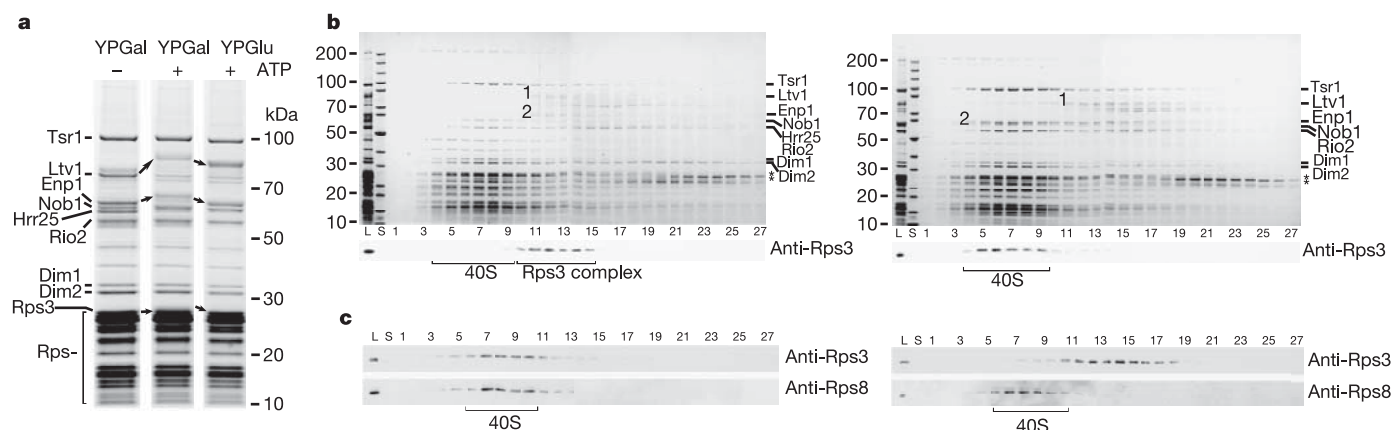


Figure 4 | Hrr25 kinase regulates 40S subunit maturation. **a**, Hrr25-dependent phosphorylation of Enp1 and Rps3. Ltv1-TAP preparations from *GAL1::HRR25* cells grown in YPGal (yeast extract, peptone, galactose) or YPGlu (yeast extract, peptone, glucose) were incubated with or without ATP and analysed by SDS-PAGE with Coomassie staining. Upward arrows, phosphorylated Ltv1, Enp1 and Rps3; downward arrows, dephosphorylated Ltv1, Enp1 and Rps3. **b**, Hrr25-dependent dissociation of Enp1 and Rps3 from pre-40S subunits. Gel-filtration profiles are shown of Ltv1-TAP

preparations from *GAL1::HRR25* cells grown for 12 h in YPGal (left) or YPGlu (right) and incubated with ATP for 30 min at 30 °C. Fractions were analysed by SDS-PAGE and Coomassie staining or western blotting (anti-Rps3 antibody). Band 1, Ltv1; band 2, Enp1. **c**, 40S subunits from Hrr25-depleted cells contain salt-extractable Rps3. Gel-filtration profiles are shown of 100 mM MgCl₂-treated mature 40S subunits isolated from *GAL1::HRR25* cells grown for 12 h in YPGal (left) or YPGlu (right). Fractions were analysed by SDS-PAGE and western blotting (anti-Rps3 and anti-Rps8 antibodies).

subunit export¹². The casein kinase I isoform Hrr25 has multiple roles in the cell²⁰ but has not been implicated in ribosome biogenesis. Notably, when pre-40S particles were purified from Hrr25-depleted cells and incubated with ATP, Enp1 and Rps3 were apparently not phosphorylated and not released from the pre-40S subunit (Fig. 4a, b). Ltv1 was partly phosphorylated in Hrr25-depleted pre-ribosomes and was dissociated from the pre-40S particle (Fig. 4a, b). In contrast, pre-40S particles isolated from Rio2-depleted cells and incubated with ATP still exhibited phosphorylated Enp1, Rps3 and Ltv1, which were dissociated at 30 °C (Supplementary Fig. S7). These data showed that the phosphorylation and subsequent dissociation of Enp1 and Rps3 from the 40S pre-ribosome is regulated by the Hrr25 protein kinase.

To determine whether Hrr25 controls Rps3 assembly *in vivo*, 40S subunits were isolated from the Hrr25-depleted cells by sucrose-gradient centrifugation and treated with a high salt concentration. Rps3 was released from the 40S subunit on incubation with 100 mM MgCl₂, whereas most ribosomal proteins, including Rps8, were not dissociated (Fig. 4c). When 40S subunits were isolated from the *GAL1::HRR25* strain grown in galactose (no Hrr25 depletion), Rps3 was not extracted by 100 mM MgCl₂ (Fig. 4c). The Hrr25-depleted cells showed defects in 18S rRNA maturation that closely resembled those seen in strains genetically depleted of Rps3 (compare Supplementary Figs S6b and S3b), strongly supporting their functional interaction *in vivo*. Depletion of Hrr25 also inhibited nuclear export of the 40S subunit (Supplementary Fig. S6c), as previously reported for Rps3 depletion¹⁸. Taken together, the data reveal that Hrr25 functions at a late step in 40S subunit biogenesis that regulates the association of Rps3 with the 40S subunit, both *in vitro* and *in vivo*.

Thus, we have uncovered a maturation step in 40S subunit biogenesis during which the binding of ribosomal protein Rps3 and the structure of the beak region of the 40S subunit are reorganized. Our data indicate that the beak RNA remains flexible while Rps3 is not tightly integrated into the 40S subunit. Hrr25-dependent phosphorylation and subsequent dephosphorylation are required for Rps3 to achieve its final association with the 40S subunit, which induces beak formation. We speculate that the incorporation of negatively charged phosphate groups into the Rps3, Enp1 and Ltv1 proteins weakens their association with the 40S subunit as a result of electrostatic repulsion. Subsequent removal of the phosphate group from Rps3 might then allow it to form a more stable association with the rRNA. The significance of the structural reorganization of the pre-40S particle remains to be determined, but a protruding, rigid beak might hinder passage through the nuclear pore complex. The nuclear pre-40S particles might retain a transport-competent structure, without a hindering beak, until transit through the nuclear pore complex. Mutations in Enp1, Ltv1 and Rps3 cause defects in the nuclear exit of pre-40S subunits, although it is unclear whether they have direct functions in export. These proteins all seem to be bound to the head region of the nascent 40S subunit, as is Rps15, another small subunit protein with a role in 40S subunit export²¹. Thus, the pre-40S head region without an exposed beak might have a key function in recruiting nuclear export factors and initiating nuclear export.

METHODS

Pre-40S ribosomal subunits were affinity-purified with TAP-tagged bait proteins Ltv1 or Rio2, as described²². Mature 40S subunits were isolated by sucrose-density centrifugation⁵ of whole cell lysates or by affinity purification of translation factor Nip1. Isolated ribosomal particles were analysed by SDS-PAGE and Coomassie staining. Associated non-ribosomal factors were identified by matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry⁵. Pre-40S and mature 40S subunits were subjected to gel-filtration chromatography under conditions of low (10 mM MgCl₂) and high (100 mM MgCl₂) salt concentration to separate subcomplexes and analyse the association state of Rps3 by western analysis²³. To compare pre-40S and mature 40S subunits at the structural level, negative-staining electron microscopy and cryoelectron microscopy with three-dimensional reconstruction were performed. An *in vitro*

assay was developed that allowed us to monitor the dependence of the maturation of pre-40S subunits on phosphorylation (addition of ATP) and dephosphorylation (phosphatase treatment). To analyse the role of *HRR25* or *RPS3* *in vivo*, pre-rRNA processing analyses^{12,24} and a fluorescence-based *in vivo* assay for nuclear export of ribosomal subunits^{25,26} were applied. For repression of *HRR25*, *RPS3* and *RIO2* gene expression in yeast, the genes were placed under the control of the *GAL1* promoter²⁷. A detailed description of the experimental procedures used is provided in Supplementary Information.

Received 22 December 2005; accepted 25 April 2006.

- Warner, J. R. Nascent ribosomes. *Cell* **107**, 133–136 (2001).
- Fatica, A. & Tollervey, D. Making ribosomes. *Curr. Opin. Cell Biol.* **14**, 313–318 (2002).
- Tschochner, H. & Hurt, E. Pre-ribosomes on the road from the nucleolus to the cytoplasm. *Trends Cell Biol.* **13**, 255–263 (2003).
- Granneman, S. & Baserga, S. J. Crosstalk in gene expression: coupling and co-regulation of rDNA transcription, pre-ribosome assembly and pre-rRNA processing. *Curr. Opin. Cell Biol.* **17**, 281–286 (2005).
- Bassler, J. *et al.* Identification of a 60S preribosomal particle that is closely linked to nuclear export. *Mol. Cell* **8**, 517–529 (2001).
- Harnpicharnchai, P. *et al.* Composition and functional characterization of yeast 66S ribosome assembly intermediates. *Mol. Cell* **8**, 505–515 (2001).
- Saveanu, C. *et al.* Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps. *EMBO J.* **20**, 6475–6484 (2001).
- Fatica, A., Cronshaw, A. D., Dlakic, M. & Tollervey, D. Ssf1p prevents premature processing of an early pre-60S ribosomal particle. *Mol. Cell* **9**, 341–351 (2002).
- Dragon, F. *et al.* A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967–970 (2002).
- Grandi, P. *et al.* 90S pre-ribosomes include the 35S pre-rRNA, the U3 snRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. *Mol. Cell* **10**, 105–115 (2002).
- Nissan, T. A., Bassler, J., Petfalski, E., Tollervey, D. & Hurt, E. 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. *EMBO J.* **21**, 5539–5547 (2002).
- Schäfer, T., Strauss, D., Petfalski, E., Tollervey, D. & Hurt, E. The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes. *EMBO J.* **22**, 1370–1380 (2003).
- Loar, J. W. *et al.* Genetic and biochemical interactions among Yar1, Ltv1 and Rps3 define novel links between environmental stress and ribosome biogenesis in *Saccharomyces cerevisiae*. *Genetics* **168**, 1877–1889 (2004).
- Frank, J., Verschoor, A. & Boublik, M. Multivariate statistical analysis of ribosome electron micrographs. L and R lateral views of the 40 S subunit from HeLa cells. *J. Mol. Biol.* **161**, 107–133 (1982).
- Frank, J., Verschoor, A. & Boublik, M. Computer averaging of electron micrographs of 40S ribosomal subunits. *Science* **214**, 1353–1355 (1981).
- Spahn, C. M. *et al.* Structure of the 80S ribosome from *Saccharomyces cerevisiae*—tRNA-ribosome and subunit-subunit interactions. *Cell* **107**, 373–386 (2001).
- Wimberly, B. T. *et al.* Structure of the 30S ribosomal subunit. *Nature* **407**, 327–339 (2000).
- Ferreira-Cerca, S., Poll, G., Gleizes, P. E., Tschochner, H. & Milkereit, P. Roles of eukaryotic ribosomal proteins in maturation and transport of pre-18S rRNA and ribosome function. *Mol. Cell* **20**, 263–275 (2005).
- Geerlings, T. H., Faber, A. W., Bister, M. D., Vos, J. C. & Raue, H. A. Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20S pre-rRNA in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **278**, 22537–22545 (2003).
- Ho, Y., Mason, S., Kobayashi, R., Hoekstra, M. & Andrews, B. Role of the casein kinase I isoform, Hrr25, and the cell cycle-regulatory transcription factor, SBF, in the transcriptional response to DNA damage in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **94**, 581–586 (1997).
- Leger-Silvestre, I. *et al.* The ribosomal protein Rps15p is required for nuclear exit of the 40S subunit precursors in yeast. *EMBO J.* **23**, 2336–2347 (2004).
- Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Sinioglou, S. *et al.* A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. *Cell* **84**, 265–275 (1996).
- Beltrame, M. & Tollervey, D. Identification and functional analysis of two U3 binding sites on yeast pre-ribosomal RNA. *EMBO J.* **11**, 1531–1542 (1992).
- Gadal, O. *et al.* Nuclear export of 60S ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Mol. Cell Biol.* **21**, 3405–3415 (2001).
- Milkereit, P. *et al.* A Noc complex specifically involved in the formation and nuclear export of ribosomal 40 S subunits. *J. Biol. Chem.* **278**, 4072–4081 (2003).

27. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **10**, 953–961 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Merker, P. Ihrig, J. Reichert, J. Pfannstiel and J. Lechner for performing mass spectrometry, and M. Seedorf and G. Dieci for the gift of antibodies. B.B. acknowledges support by a grant from EU-NOE (3D-Repertoire). E.H. is recipient of grants from the Deutsche Forschungsgemeinschaft and Fonds der Chemischen Industrie. E.P. and D.T. were supported by the Wellcome Trust.

Author Contributions Experiments were designed and data were analysed and interpreted by T.S. and E.H. Strain constructions, DNA recombinant work,

fluorescence microscopy and biochemical analyses (affinity purification, gel filtration, sucrose gradient centrifugation and *in vitro* assays) were performed by T.S. Negative-staining electron microscopy was conducted by B.M. and U.A., and cryoelectron microscopy and three-dimensional reconstruction by B.B. E.P. and D.T. performed rRNA processing analyses. The manuscript was written by T.S. and E.H. All authors discussed the results and commented on the manuscript.

Author Information Three-dimensional reconstructions of mature 40S and pre-40S ribosomal subunits have been deposited in the EMBL-EBI Molecular Structure Database (<http://www.ebi.ac.uk/msd/>) and can be retrieved under accession numbers EMD-1211 and EMD-1212. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.H. (cg5@ix.urz.uni-heidelberg.de).

LETTERS

Computational redesign of endonuclease DNA binding and cleavage specificity

Justin Ashworth¹, James J. Havranek¹, Carlos M. Duarte¹, Django Sussman³, Raymond J. Monnat Jr², Barry L. Stoddard³ & David Baker¹

The reprogramming of DNA-binding specificity is an important challenge for computational protein design that tests current understanding of protein–DNA recognition, and has considerable practical relevance for biotechnology and medicine^{1–6}. Here we describe the computational redesign of the cleavage specificity of the intron-encoded homing endonuclease I-MsoI⁷ using a physically realistic atomic-level forcefield^{8,9}. Using an *in silico* screen, we identified single base-pair substitutions predicted to disrupt binding by the wild-type enzyme, and then optimized the identities and conformations of clusters of amino acids around each of these unfavourable substitutions using Monte Carlo sampling¹⁰. A redesigned enzyme that was predicted to display altered target site specificity, while maintaining wild-type binding affinity, was experimentally characterized. The redesigned enzyme binds and cleaves the redesigned recognition site ~10,000 times more effectively than does the wild-type enzyme, with a level of target discrimination comparable to the original endonuclease. Determination of the structure of the redesigned nuclease–recognition site complex by X-ray crystallography confirms the accuracy of the computationally predicted interface. These results suggest that computational protein design methods can have an important role in the creation of novel highly specific endonucleases for gene therapy and other applications.

The nucleotide sequence specificity of DNA-binding proteins can not be deduced directly from amino acid sequence because the packing, hydrogen-bonding and electrostatic interactions responsible for nucleotide-specific recognition are dependent on the three-dimensional structure of the protein–DNA complex^{11,12}. While a number of canonical amino acid–nucleotide interaction motifs are observed in protein–DNA interfaces¹³, they are not in general predictable from sequence information alone. Hence, in place of a simple ‘recognition code’, an atomic-level model of the protein–DNA interface is likely to be necessary to fully capture the basis of recognition specificity. To understand the specificity of naturally occurring DNA-binding proteins, and to design new specificities, we have developed a computational model that explicitly treats the packing, hydrogen-bonding, solvation and electrostatic interactions that underlie protein–DNA interactions^{9,14}. Here we describe the use of this model to redesign the specificity of the I-MsoI homing endonuclease.

I-MsoI, which belongs to the LAGLIDADG family of homing endonucleases, is a 170-residue homodimeric enzyme that cleaves long target sites (20–24 base pairs (bp)) with considerable specificity^{7,15,16}. The homing endonucleases provide an excellent model system for understanding protein–DNA interaction specificity, as well as starting points for engineering of novel specificities for targeted genomics applications, including gene therapy^{4,5}. Crystal

structures of the enzymes bound to their recognition sites reveal a rich assortment of side-chain–nucleotide contacts within the DNA major groove, which provide many possibilities for the redesign of specificity^{16,17}.

We first tested our model of the DNA–protein interface by repacking the sidechains at the native I-MsoI interface. Nearly all protein–DNA contacts and most sidechain dihedral angles in the crystal structure were reproduced (Supplementary Fig. S1, Supplementary Table S1). To benchmark the sampling and evaluation of alternative amino acid identities required for protein design, clusters of amino acids were redesigned around each native base pair. All direct hydrogen-bonding contacts between protein and nucleotide bases present in the wild-type complex were preserved, showing that the model captures the important aspects of the naturally occurring interface. To redesign I-MsoI DNA cleavage specificity, we began by screening *in silico* for base changes predicted to disrupt binding by the wild-type enzyme (Supplementary Figs S2, S3). The amino acids in the vicinity of each of the base-pair substitutions predicted to disrupt binding were then redesigned, and the designs were ranked on the basis of the predicted affinity of the designed protein for the new site, and the predicted decrease in affinity of the native enzyme for the new site (Supplementary Fig. S4).

The design with the largest predicted change in specificity consisted of the base-pair substitution –6C·G to –6G·C in the ‘left’ DNA half-site, and a similar change in the symmetry-related ‘right’ half-site from +6A·T to +6C·G (Supplementary Fig. S5; numbers are the distance in base pairs from the centre of the recognition site). At both positions in the wild-type complex, the key interactions of the base pair are a hydrogen bond of the purine ring with Lys 28 and a water-mediated contact with Thr 83 (Fig. 1a). Converting either base pair to a G·C is predicted to disrupt binding (+3.2 kcal) by the loss of the direct hydrogen-bonding interaction and the resulting desolvation of Lys 28 (Fig. 1b).

Compensation of the base-pair substitution by redesign of the surrounding amino acids yielded the low energy solution K28L, T83R (–4.2 kcal versus wild type). As shown in Fig. 1d, Arg 83 is predicted to make two hydrogen bonds to the guanine nucleotide of the introduced G·C base pair, a sidechain–base interaction motif important in naturally occurring protein–DNA interfaces^{13,18}. The Leu 28 mutation decreases the binding energy of the designed enzyme to wild-type DNA by eliminating a purine-specific hydrogen-bond. Furthermore, Leu 28 in the designed enzyme makes a favourable non-polar packing contact with the C5 of cytosine of the designed DNA (Fig. 1d). This leucine may also contribute to specificity against a purine at this position by unfavourably burying polar surface area of nitrogen N7 (Fig. 1c). The predicted binding

¹Howard Hughes Medical Institute and Department of Biochemistry, ²Departments of Pathology and Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ³Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue, Seattle, Washington 98109, USA.

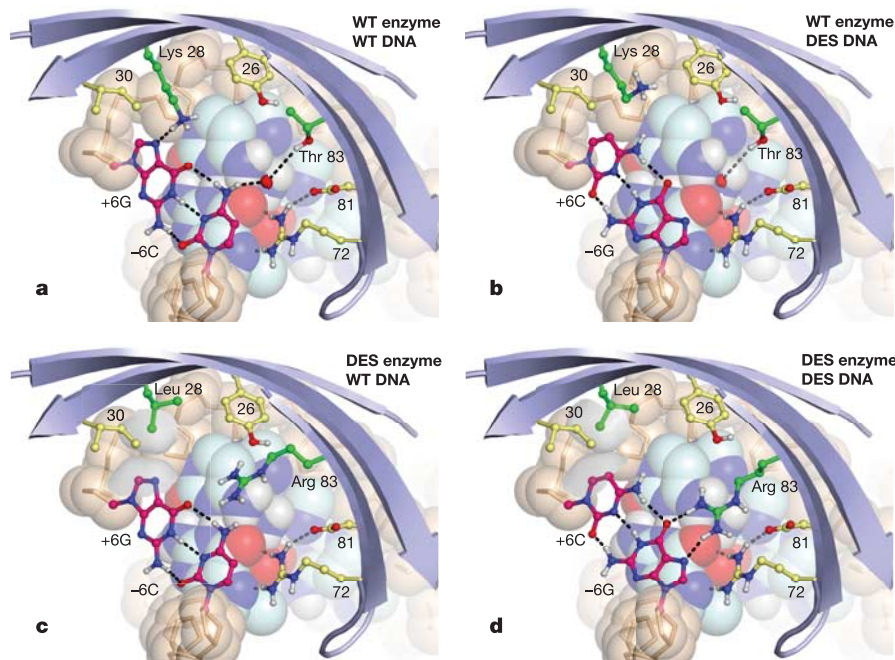


Figure 1 | Comparison of the predicted interactions in cognate and non-cognate binding complexes, illustrating the designed specificity switch. **a**, Wild-type *I-MsoI*, $-6C\cdot G$ (wild type). A water molecule present in the original structure¹⁶ is shown. **b**, Wild-type *I-MsoI*, $-6G\cdot C$. **c**, *I-MsoI*-K28L/T83R, $-6C\cdot G$. **d**, *I-MsoI*-K28L/T83R, $-6G\cdot C$. In parts **c** and **d**, the van der

Waals surfaces of Leu 28 and $+6C$ are shown in grey. Figures were generated using the molecular graphics program PyMOL (Delano Scientific). WT, wild type; DES, designed; blue strands, protein backbone; beige spheres and sticks, DNA backbone; other spheres, constant nucleotides; dashed lines, hydrogen bonds.

energies of the cognate and non-cognate complexes are given in Table 1. Competitive *in vitro* cleavage assays^{15,19} were performed to assess the specificity of the wild-type and designed enzymes by directly comparing the relative activity of the enzyme on each recognition site. Specific site cleavage, in the presence of both the cognate and non-cognate sites, in addition to 8.3 kilobases (kb) of non-specific vector sequence, is independently observed as the conversion of the linearized plasmid band into two smaller fragments of unique size. As is evident in the upper panel in Fig. 2, 100 nM wild-type *I-MsoI* cleaves a substantial fraction of the wild-type target site, but little or no cleavage of the designed site occurs at concentrations as high as 6.4 μ M. In contrast, cleavage of the designed site by the designed enzyme (Fig. 2, lower panel) is observed at an enzyme concentration of 200 nM, whereas the original site is cleaved only at enzyme concentrations of 3.2 μ M and above. To determine the specificity of binding independent of catalysis, gel electrophoretic mobility shift assays of cognate and non-cognate DNA–protein complexes were performed, and a similar switch in specificity was observed (Table 2). The shift in sequence specificity suggested by these results (at least 4,000-fold by competitive cleavage assay, and $\sim$ 13,000-fold by gel shift assay) is considerably greater than that shown to be required for changes in phenotype *in vivo* gene elimination assays using altered homing endonucleases^{19,20}. To verify the atomic-level accuracy of the computationally predicted design, the crystallographic structure of the complex was determined to 2.0 Å resolution (Supplementary Fig. S6). A difference

map showing the electron density around the redesigned amino acids superimposes well with the predicted structure, confirming the accuracy of the design (Fig. 3). A water molecule is also evident at the site, but it does not appear to contribute to nucleotide-binding specificity. Our results represent a significant advance in the redesign of protein–DNA interaction specificity, and demonstrate the efficacy of explicit, atomic-level modelling of the protein–DNA interface for the re-engineering of specificity. While the subject of this work is a homing endonuclease, the method should be generalizable to any protein–DNA interface redesign problem: for example, the reprogramming of transcription factor binding specificity. The

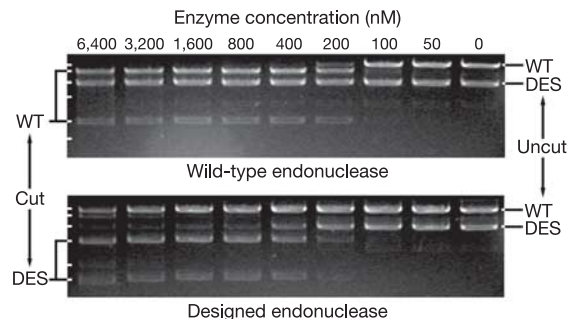


Figure 2 | Switch in nuclease cleavage specificity. Equimolar amounts of linearized plasmid DNAs containing wild-type (WT) or designed (DES) *I-MsoI* cleavage sites were digested by serial dilutions of wild-type or designed *I-MsoI* endonuclease, and analysed by gel electrophoresis. The switch in sequence specificity is defined as (wild type vs. DES/wild type vs. WT) $\times$ (designed vs. WT/designed vs. DES), where quantities in parentheses indicate the lowest enzyme concentration at which significant cleavage of the site is observed. Here, the wild-type enzyme favours the WT site by $>2^7$ -fold, the designed enzyme favours the DES site by $\sim 2^5$ -fold, and hence the specificity switch is greater than $2^7 \times 2^5$ ($>4,000$ -fold).

Table 1 Predicted binding energies*		
Target sites	<i>I-MsoI</i>	<i>I-MsoI</i> -K28L-T83R
$-6C\cdot G, +6A\cdot T$	0.0 (0.0)	+1.6 (+1.58)
$-6G\cdot C, +6C\cdot G$	+3.2 (+2.34)	−4.2 (−1.13)

*Relative binding energies of designed cognate and non-cognate complexes were computed as (energy of complex)–(energy of isolated DNA + isolated protein), with the value for the wild-type complex subtracted to facilitate comparison. The corresponding values for the hydrogen-bonding contribution to the total energy are shown in parentheses.

Table 2 | Experimental binding affinities*

Target sites	Protein	
	I-MsoI	I-MsoI-K28L-T83R
–6C•G, +6A•T	61 ± 15 nM	6.1 ± 1.3 μM
–6G•C, +6C•G	>25 μM	192 ± 30 nM

* Binding affinities for wild-type and designed I-MsoI as determined by gel electrophoresis shift. Errors represent 67% confidence intervals.

computational approach described here can be improved further by modelling protein and DNA backbone flexibility and water-mediated interactions. Remaining inaccuracies in the potential function can potentially be compensated by focused exploration around promising computational designs using experimental selection methodologies.

The engineering of artificial gene-specific reagents from naturally occurring DNA-binding proteins is of great interest for a variety of targeted genetic applications. Site-specific zinc-finger nucleases (ZFNs), generated as chimaeras of non-specific endonuclease domains fused to zinc-finger domains, can stimulate gene-specific homologous recombination^{2,3} and have recently been shown to promote the repair of a disease-associated mutation in cultured cells⁶. The homing endonucleases are an alternative molecular system for the creation of gene-specific DNA cleavage enzymes⁴ that have also been shown to elicit gene repair by homologous recombination in murine hepatocytes²¹. These proteins are inherently more site-specific in their DNA cleavage activities than are ZFNs, and have the added advantage that site binding and cleavage are tightly coupled in the same protein domain, perhaps minimizing off-target cleavage. The redesign of existing homing endonuclease proteins to confer novel DNA target specificities remains challenging using methods of directed evolution^{5,20,22}. The use and refinement of the computational modelling and design strategies described here should facilitate such efforts, and allow us to approach our long-term goal of designing novel proteins able to recognize and cleave any desired DNA site with high specificity for targeted genomics applications.

METHODS

Computational design. Models of every single base-pair substitution in the I-MsoI–target site complex were generated, and a Monte Carlo search procedure was used to sample alternative conformations and identities of the surrounding amino acid side chains. Protein positions were repacked or redesigned if at least one arginine rotamer placed at the position could make contact to the substituted DNA. Side-chain rotamer conformations were taken from the Dunbrack library²³, and supplemented with extra rotamers generated by varying χ_1 and χ_2 independently by plus or minus one standard deviation of the principal distribution for the rotamer. For still finer sampling of the conformations of long polar side chains, additional rotamers were generated by similarly perturbing χ_3 and χ_4 ; these were retained for the full combinatorial search if they had favourable hydrogen-bonding energies with the DNA bases. The physically realistic all atom force field used to guide the Monte Carlo search is composed of a Lennard–Jones-based treatment of packing, an orientation dependent hydrogen-bonding potential, a generalized Born-based treatment of electrostatic solvation energy²⁴, a PDB-derived side-chain torsional potential, and amino acid dependent reference energies which represent average residue energies in the unbound and unfolded state. The lowest energy structures identified in the Monte Carlo search were further optimized by continuous minimization of side-chain and DNA torsion angles using the Powell method^{14,25}. The *in silico* protein–DNA complexes produced by sequence design followed by minimization were ranked on the basis of the predicted affinity of the designed enzyme for the new site, and the predicted loss in affinity of the wild-type enzyme for the new site (Supplementary Fig. S4).

Experiments. Full details of the experimental methods are given in Supplementary Information; a summary is given below.

Expression and purification of I-MsoI was performed as previously described¹⁶. Substrates for competitive cleavage were as follows: oligonucleotide duplexes corresponding to I-MsoI_{WT} (5′-GCAGAACGTCGTGAGACAGTTCCG-3′; bold font indicates positions that were changed in the design) and I-MsoI_{DES} (5′-GCAGAAGGTCGTGAGACCGTTCCG-3′) cleavage sites were incorporated

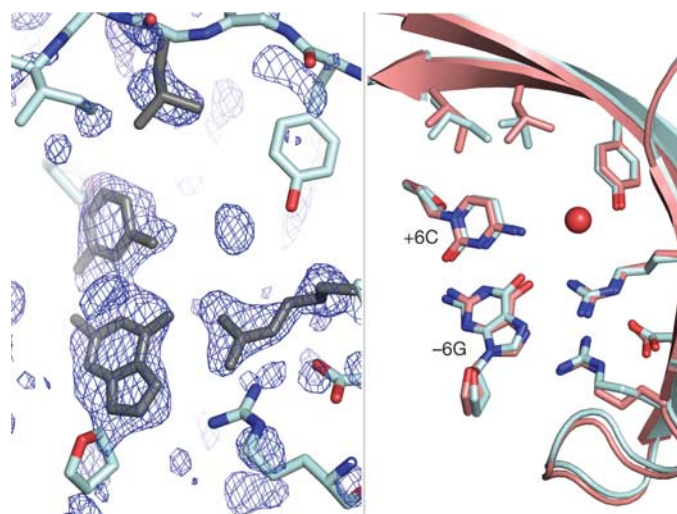


Figure 3 | Crystal structure of the designed enzyme–DNA complex. Left, $F_o - F_c$ electron-density map of the redesigned region calculated from a refinement model lacking the redesigned side chains and bases (cyan). The computational design model (grey) fits well into the unassigned density (blue mesh, +2.2 σ). Right, superposition of the design model (salmon) and the refined crystal structure (cyan) confirms the accuracy of the design. A new coordinated water molecule (red sphere) is also apparent.

into plasmid vectors of sizes 5.4 kb (wild-type site) and 2.9 kb (designed site). To facilitate product identification, the plasmid substrates were linearized by restriction-enzyme digestion before use in cleavage assays.

Serial twofold dilutions of wild-type and designed protein were added to reaction mixtures containing 50 nM of both linearized cleavage constructs in the presence of magnesium. Reactions proceeded for 1 h at 37 °C. Electrophoretic mobility shift assays to determine binding constants of the cognate and non-cognate complexes were carried out in the presence of calcium, using 5′-radiolabelled DNA duplexes and non-specific competitor DNA. Crystals of I-MsoI-K28L-T83R bound to designed DNA duplex formed in previously described conditions¹⁶. Data were collected at the Advanced Light Source. The structure was solved by molecular replacement using the original I-MsoI structure, and refined at 2.0 Å resolution using CNS²⁶. The final refinement statistics (Supplementary Table S2) for I-MsoI-K28L-T83R were R_{work}/R_{free} = 0.229/0.271.

Received 9 January; accepted 21 April 2006.

- Uil, T. G., Haisma, H. J. & Rots, M. G. Therapeutic modulation of endogenous gene function by agents with designed DNA-sequence specificities. *Nucleic Acids Res.* **31**, 6064–6078 (2003).
- Bibikova, M. *et al.* Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.* **21**, 289–297 (2001).
- Porteus, M. H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).
- Wickelgren, I. Molecular biology. Spinning junk into gold. *Science* **300**, 1646–1649 (2003).
- Stoddard, B. L. Homing endonuclease structure and function. *Q. Rev. Biophys.* **38**, 1–47 (2005).
- Urnov, F. D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
- Lucas, P., Otis, C., Mercier, J. P., Turmel, M. & Lemieux, C. Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases. *Nucleic Acids Res.* **29**, 960–969 (2001).
- Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93 (2004).
- Havranek, J. J., Duarte, C. M. & Baker, D. A simple physical model for the prediction and design of protein–DNA interactions. *J. Mol. Biol.* **344**, 59–70 (2004).
- Voigt, C. A., Gordon, D. B. & Mayo, S. L. Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *J. Mol. Biol.* **299**, 789–803 (2000).
- Kono, H. & Sarai, A. Structure-based prediction of DNA target sites by regulatory proteins. *Proteins* **35**, 114–131 (1999).
- Pabo, C. O. & Neukludova, L. Geometric analysis and comparison of protein–DNA interfaces: why is there no simple code for recognition? *J. Mol. Biol.* **301**, 597–624 (2000).

13. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. Amino acid-base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Res.* **29**, 2860–2874 (2001).
14. Morozov, A. V., Havranek, J. J., Baker, D. & Siggia, E. D. Protein–DNA binding specificity predictions with structural models. *Nucleic Acids Res.* **33**, 5781–5798 (2005).
15. Seligman, L. M. *et al.* Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.* **30**, 3870–3879 (2002).
16. Chevalier, B., Turmel, M., Lemieux, C., Monnat, R. J. Jr & Stoddard, B. L. Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J. Mol. Biol.* **329**, 253–269 (2003).
17. Heath, P. J., Stephens, K. M., Monnat, R. J. Jr & Stoddard, B. L. The structure of I-CreI, a group I intron-encoded homing endonuclease. *Nature Struct. Biol.* **4**, 468–476 (1997).
18. Seeman, N. C., Rosenberg, J. M. & Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl Acad. Sci. USA* **73**, 804–808 (1976).
19. Sussman, D. *et al.* Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J. Mol. Biol.* **342**, 31–41 (2004).
20. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.* **128**, 2477–2484 (2006).
21. Gouble, A. *et al.* Efficient *in toto* targeted recombination in mouse liver by meganuclease-induced double-strand break. *J. Gene Med.* published online 13 February 2006 (doi:10.1002/jgm.879) (2006).
22. Arnould, S. *et al.* Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.* **355**, 443–458 (2006).
23. Dunbrack, R. L. Jr & Cohen, F. E. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661–1681 (1997).
24. Onufriev, A., Bashford, S. D. & Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized Born model. *Proteins* **55**, 383–394 (2004).
25. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge Univ. Press, New York, 1992).
26. Brunger, A. T. *et al.* Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. L. Eklund for assistance with binding assays, and B. W. Shen for assistance with data collection and refinement. This work was supported by fellowships from the Jane Coffin Childs Memorial Fund (J.J.H.), the National Science Foundation (C.M.D.), and grants from the National Institute of Health (R.J.M. and B.L.S.), the Howard Hughes Medical Institute (D.B.), and the Gates Foundation Grand Challenges Program (B.L.S., D.B., R.J.M.).

Author Contributions J.J.H. and C.M.D. developed the original protein–DNA interface design methods and code. J.A. made further code and method developments, generated and assessed the computational predictions, and performed mutagenesis, biochemical characterization, and crystallization. D.S. collected and processed the crystallographic data, and aided in protein purification and structure refinement.

Author Information The atomic coordinates of the redesigned I-MsoI endonuclease bound to its cognate DNA have been deposited in the Protein Data Bank with the accession number 2FLD. Reprints and permissions information are available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.A. (ashwortj@u.washington.edu) or D.B. (dabaker@u.washington.edu).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Some say the movies are a mirror — but in the case of scientists, that reflection is in the funhouse style, distorting people into grotesque caricatures. I recently encountered a photo essay on scientists in movies, which offered more than a history lesson in pop culture. Its subtext showed how the perception of scientists is changing, perhaps offering a way to a better image.

The first 'mad' scientist appeared in the 1926 Fritz Lang film *Metropolis*. Rotwang sported unruly hair, a disabled hand and a maniacal approach to research. The character reflected myths about science — that it was the result of obsessive individuals who didn't care about science's effects on society. This perception isn't new in art: the principal theme in Goethe's *Faust* also shows the dangers of pursuing bookish knowledge.

But movies have a way of overplaying that danger. In Mary Shelley's novel *Frankenstein*, the monster was described as "beautiful" — a positive result in aesthetic terms for science. But movie depictions of the creature have been uniformly menacing, with the malevolent side of the monster overtaking the gentle.

More modern movies still show scientists as clueless about the implications of their research. *The Nutty Professor* turned comic actor Eddie Murphy into an overweight and out-of-control Mr Hyde. And yet some scientists are finally being displayed as maverick heroes, fighting to get their ideas out when the establishment is trying to contain them. For instance, Jodie Foster, the heroine of *Contact*, defies government forces to get in touch with extraterrestrials.

This bodes well for the future, as many scientific ideas are now being distorted by these establishment forces — such as stem cells, evolution and global warming. If real-life scientists continue to rage against the establishment machine and tell the public how science can help solve problems, rather than create them, they are more likely to be depicted as benevolent characters on the big screen. Perhaps in the process, they will encourage more people to consider scientific careers, and attract the funding to pay for them.

Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe

Editor: Paul Smaglik

Assistant Editor: Gene Russo

European Head Office, London

The Macmillan Building,

4 Crinan Street

London N1 9XW, UK

Tel: +44 (0) 20 7843 4961

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

European Sales Manager:

Andy Douglas (4975)

Natureevents: Matt Betteridge (4015)

UK/RoW/Ireland:

Nils Moeller (4953)

Scandinavia/Spain/Portugal:

Claudia Paulsen Young (4973)

France/Switzerland/Belgium/Italy:

Maria Ceraolo (4994)

Germany/Austria/The Netherlands:

Reya Silao (4970)

Business Development Manager:

Amelie Pequignot (4974)

Advertising Production Manager:

Stephen Russell

To send materials use London

address above.

Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Catherine Alexander

European Satellite Office

Germany:

Patrick Phelan

Tel: +49 89 54 90 57 11

Fax: +49 89 54 90 57 20

e-mail: p.phelan@nature.com

US Head Office, New York

75 Varick Street,

9th Floor, New York,

NY 10013-1917

Tel: +1 800 989 7718

Fax: +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

Chiyoda Building,

2-37 Ichigayatamachi,

Shinjuku-ku,

Tokyo 162-0843

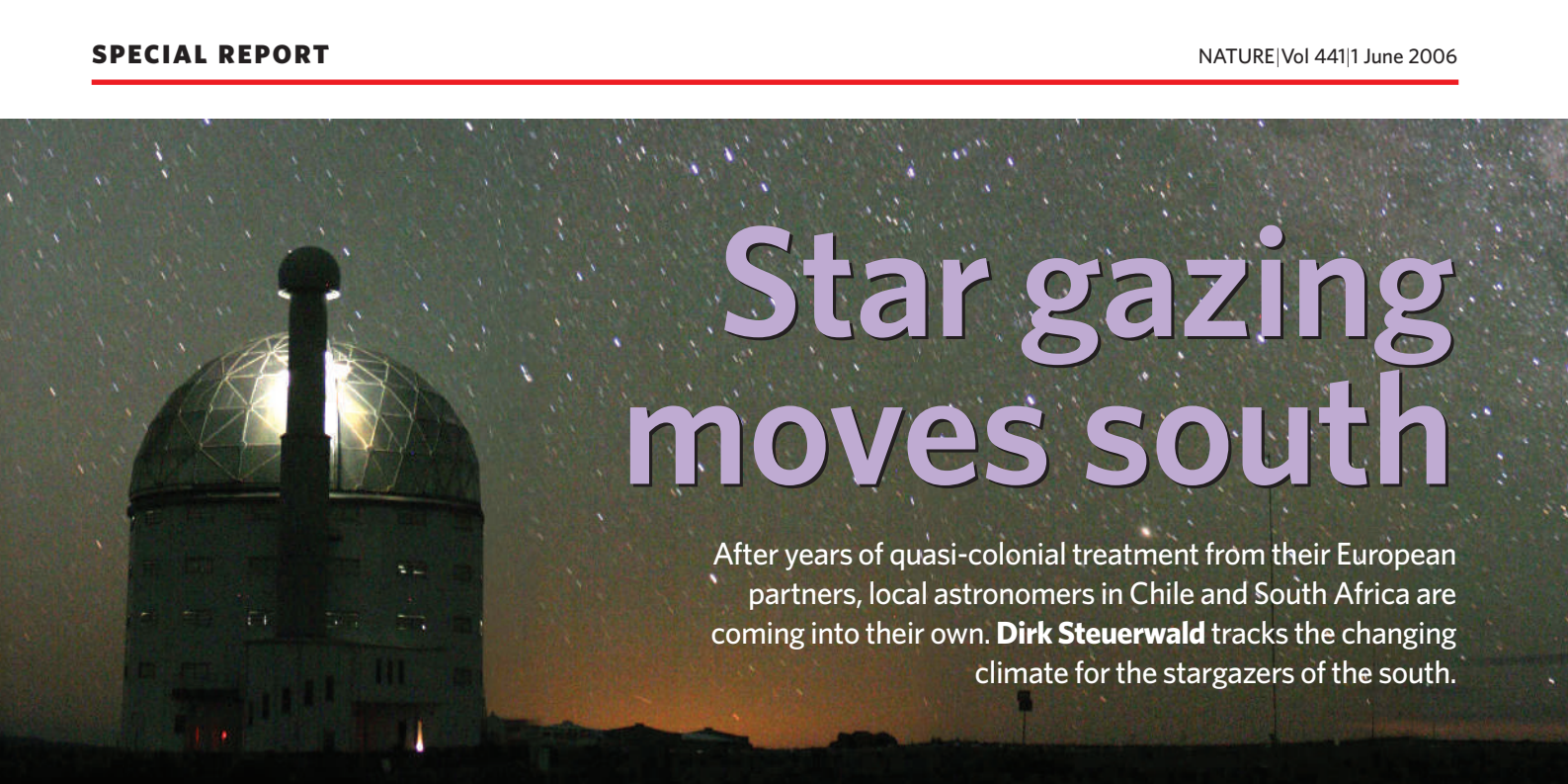
Tel: +81 3 3267 8751

Fax: +81 3 3267 8746

Asia-Pacific Sales Director:

Ayako Watanabe

e-mail: a.watanabe@natureasia.com



Star gazing moves south

After years of quasi-colonial treatment from their European partners, local astronomers in Chile and South Africa are coming into their own. **Dirk Steuerwald** tracks the changing climate for the stargazers of the south.

In 1820, the Reverend Fearon Fallows, a British astronomer and clergyman, set up what would become South Africa's first permanent observatory. Fallows's mission, sparked by imprecise celestial charts causing too many shipwrecks, marked the beginning of the 'colonial' period of astronomy, the impact of which still reverberates today.

Historically, many of the observatories in the southern hemisphere have resembled high-tech islands in scientifically underdeveloped environments. But this is beginning to change. The clear night skies and dry atmosphere in this part of the world have attracted a growing number of professional stargazers, as well as engineers and technicians, from Europe and North America. But in recent years, an increasing number of local scientists have started to participate in astronomy at modern observatories and universities in their home countries. The job market for local astronomers in the southern hemisphere is expanding rapidly, notably in Chile, which hosts some of the finest and most powerful telescopes in the world.

Not only has the number and quality of astronomers increased, but associated disciplines have also been benefiting from the boom. At Chilean facilities, for example, more than 80% of engineering, technical-support and civil-development staff are recruited from the local job market, according to Felix Mirabel, head of the European Southern Observatory (ESO) office for science in Santiago.

Boom time

Thanks to its ideal geographical location and atmospheric conditions, Chile has become an astronomy hot spot. The country will cement its standing in the field when the Atacama Large Millimetre Array (ALMA), a US\$1-billion facility is installed 5,000 metres above sea level in the Atacama desert. This is expected to begin operations in 2011.

Adding to this, in May, an international group of astronomers selected Cerro Pachón, a 2,682-metre mountain peak in northern Chile, as the site for the proposed Large Synoptic Survey Telescope. Funding

uncertainties notwithstanding, the telescope is scheduled to see first light in 2012.

The first big international observatory, the Cerro Tololo Inter-American Observatory (CTIO), arrived in Chile 40 years ago, but as a domestic discipline Chilean astronomy is relatively new. Twenty years ago, most of the scientific, technical and engineering staff working at Chilean observatories were foreigners. But the small community became increasingly unhappy with the quasi-colonial interaction they had with European collaborators.

In the late 1960s, an agreement between Chile and the US National Optical Astronomy Observatory — America's national centre for ground-based optical astronomy — granted Chilean astronomers 10% of observatory viewing time. However, the country still had no such agreement with ESO. In the early 1990s, Chilean astronomers, such as Maria Teresa Ruiz, demanded that Chileans get more in return for providing ground and infrastructure (*Nature* 363, 384; 1993).

The complaints were not in vain. A 1996 agreement with ESO guaranteed 10% observing time for the locals. "The situation has changed completely in the past few years," says Ruiz, long-time director of astronomy at the University of Chile in Santiago and chair of the Chilean Telescope Time Allocation Committee. "Everybody is happy and trusts each other."

Some critics say, however, that the relatively small Chilean astronomical community is not yet able to use its observing time effectively. Competition for observing time is much less pronounced in Chile than in Europe and the United States, where typically only one out of five applications is approved. In Chile, one of every two is approved. "We're still a small community," says Mónica Rubio, a professor of astronomy at the University of Chile and president of the Chilean Astronomical Society, noting that researchers are focused on projects with the biggest potential impact. Chilean astronomers are logging significant hours on the large but not the small telescopes, she says.

"We need to train more young scientists, and enable



Astronomers Patricia Whitelock and Felix Mirabel think investment in education is key to their countries' long-term success in astronomy.

them to make the best possible use of observation time,” Ruiz says. Because the infrastructure already exists, Chile can afford to focus resources on education. “It makes a lot of sense for us to invest in brains rather than in telescopes,” says Ruiz. “We can achieve a lot with only a small amount of money.”

This strategy is already being employed. In 1995, Chilean universities had only 22 full-time astronomers. By 2005, the number had almost tripled to 64, of which 42 are full professors. Over the past ten years, the number of postgraduate students enrolled in astronomy has increased eightfold, says Mirabel. To strengthen training opportunities in the field, three Chilean universities have recently established PhD programmes for astronomers. And the astronomy department at the University of Chile is developing its own millimetre-wave laboratory in order to refine receivers for the ALMA telescope in collaboration with scientists in Europe.

All aboard

Because of racism and political instability, South Africa has long been unpopular as a host country for the international astronomy community.

That changed when apartheid was abolished in 1994. South Africa now hosts the largest single optical telescope in the southern hemisphere. Called the South African Large Telescope (SALT), it began operation in September 2005. SALT was built in collaboration with the United States, Poland, Germany, Britain and New Zealand. But unlike most telescopes in Chile, the facility is not merely a western export of scientific infrastructure. Plans for its development came from South African scientists who were eager to keep pace with the international astronomy community.

SALT will achieve cutting-edge astronomical observations at a near-perfect site at Sutherland in the Northern Cape Province, according to Patricia Whitelock, head of the astronomy division at the South African Astronomical Observatory (SAAO), in Cape Town. South African scientists should be able to effectively use their 35% share of the total SALT observation time available, she says. Whitelock notes that two large surveys, one galactic and one cosmological, have been proposed, each of which alone would use the country's share of observation time from SALT for more than a year.

She adds, however, that South Africa lacks a sufficient number of well-trained postdocs and PhD students to analyse the expected wealth of observations. At present, the South African community consists of

only 65 astronomers spread across the country, which makes teaching on a high academic level difficult. The SAAO's National Astrophysics and Space Science Programme, set up in 2003, has been tasked with producing the needed researchers. It will target in particular the disadvantaged communities of women and black scientists.

The programme, funded jointly by private foundations and the South African National Research Foundation, is based at the University of Cape Town. Lecturing staff hail from the entire South African astronomy and space-physics community. Students lacking funds receive generous bursaries, a crucial factor given that poverty is commonplace. By the end of 2006, 50 honours students and 20 MSc students will have graduated. Before the programme, South Africa had no more than one honours graduate per year.

But the programme has recently run to a deficit. The government has promised additional support but hasn't yet allocated the money, leading to concerns over the programme's continuation. “It would be a disaster for South Africa if we simply end up as technicians running telescopes and leave our international partners to do the science,” says Whitelock.

World leaders

But leading scientists at ESO predict that Chile and South Africa will further strengthen their role in international astronomy. “Chile has very good astronomers and manages to use its time pretty efficiently,” says Bruno Leibundgut, head of ESO's office for science in Garching, Germany.

The example of Spain, he says, illustrates how much a host country can benefit from foreign investment in astronomy facilities. Astronomy in Spain blossomed after its first international observatory was built at Calar Alto near Almería in the 1970s. Back then the tiny Spanish community consisted of no more than 20 astronomers. Today, Spain has around 500 astronomers, each of whom publishes one paper per year on average; Spanish astronomers account for 6% of the publication output in astronomy world-wide.

Spain also has acquired the expertise and financial potential necessary to build its own facilities. By the end of the year, the Gran Telescopio Canarias, a large optical and infrared telescope spanning ten metres across, will begin operation on the Canary Island of La Palma (*Nature* 435, 140–142; 2005). The telescope is a mostly Spanish-built project with only a 10% share held by the United States and Mexico.

For astronomers, Spain, which signed a pre-agreement to join ESO in February, has an excellent job market. “We are still short of astronomers,” says Xavier Barcons, a research professor at the Institute of Physics of Cantabria, who manages Spain's astronomy and astrophysics programme.

But countries such as Chile and South Africa still face major challenges in a field that's changing rapidly, according to Rubio. Researchers must store and process heaps of data streaming in from survey telescopes. Data archiving and mining capabilities are increasingly important. Whereas the United States and Europe have started to build databases to deal with the data deluge, Chile has yet to cultivate the necessary technical expertise or infrastructure. “It's a big challenge for Chilean astronomy,” she says.

Dirk Steuerwald is an intern in Nature's Munich office.

ESO



Testing is under way in preparation for the Atacama Large Millimetre Array, a powerful telescope that will be hosted in Chile.



THE NATIONAL INSTITUTES OF HEALTH WWW.NIH.GOV

National Institute on Aging Intramural Research Program

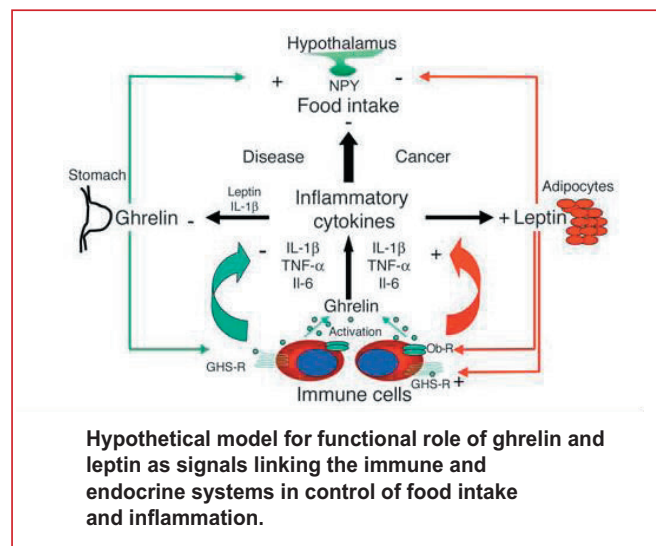
The Intramural Research Program (IRP) in the National Institute on Aging (NIA) comprises 11 scientific laboratories, a clinical branch, a research resources support branch and 2 sections. Most research is conducted at the program's headquarters, the Gerontology Research Center, in Baltimore, Maryland; however, two laboratories and one section are located in Bethesda, Maryland. The research program includes the scientific disciplines of biochemistry, cell and molecular biology, structural biology, genetics, immunology, neurogenetics, behavioral sciences (psychology, cognition, and psychophysiology), epidemiology, statistics, and clinical research and the medical disciplines of neurobiology, immunology, endocrinology, cardiology, rheumatology, hematology, oncology, and gerontology. Medical problems associated with aging are pursued in depth using the tools of modern laboratory and clinical research. Focused research seeks to understand the age-related changes in physiology and the ability to adapt to environmental stress. This understanding is used to develop insight about the pathophysiology of age-related diseases, such as Alzheimer's disease, Parkinson's disease, stroke, atherosclerosis, osteoarthritis, diabetes, and cancer. The program seeks to understand the changes associated with healthy aging, to define the criteria for evaluating when a change becomes pathologic, and to utilize this knowledge translationally to provide improved prevention strategies, diagnostic and therapeutic options.

Scientific Advances and Accomplishments: A Diverse Research Portfolio

Research performed by NIA-IRP scientists has significantly contributed to our understanding and treatment of many of age-related diseases. For example, NIA scientists discovered that Paclitaxel (Taxol), a drug used to treat cancer, markedly attenuates vascular restenosis, when used to coat stents inserted during angioplasty. More than 750,000 angioplasty procedures are performed in the United States every year and most require the placement of stents to support the artery wall. The clinical application of this NIA finding, has dramatically improved both the short and long-term success rates of this form of treatment for cardiovascular disease. NIA scientists have also been at the forefront in developing novel treatment modalities for type 2 diabetes mellitus, a condition that now affects approximately 21% of all Americans over age 60 and a rapidly increasing percentage of the population in the adolescent and young adult age groups.

The science pursued in the program is diverse and yet focused on understanding how basic biological, biochemical, and biophysical processes evolve to culminate in what we know as the aging phenotype. Recent laboratory advances have included:

- The discovery of FANCB, an X-chromosome gene that protects against Fanconi Anemia (FA), a rare genetic disease characterized by congenital defects, bone marrow failure, and cancer susceptibility.
- The identification of an inherited form of Parkinson's disease that is caused by the triplication of the PARK1 gene, which results in the overproduction of the protein alpha-synuclein and the formation of a particular protein aggregate found in the brains of Parkinson's patients.
- The finding that ghrelin, previously described as a potent circulating orexigen, controlling the metabolic axis, is also expressed in T lymphocytes and monocytes and exerts a potent anti-inflammatory effect. This finding couples the immune system to metabolism and supports the potential use of ghrelin and agonists to its receptor as potential therapeutic targets in the



management of inflammatory diseases.

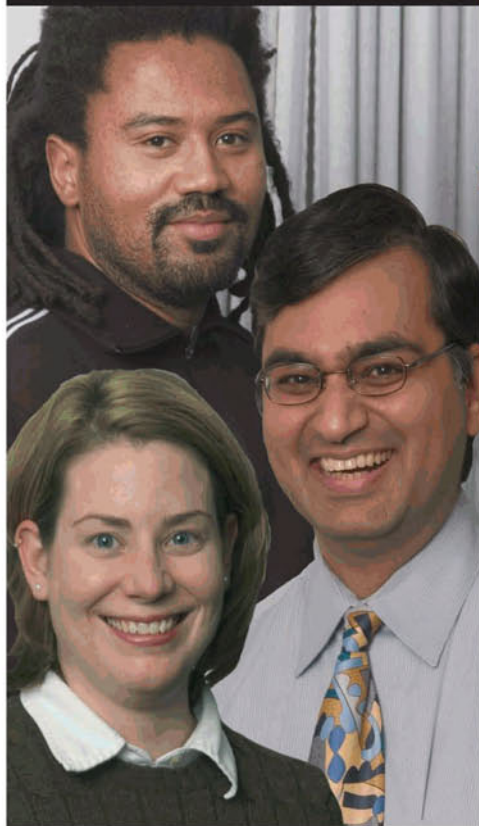
- The observation that socioeconomic status affects health outcomes at all stages of life. Relating childhood socioeconomic environment to midlife functional status provides a life course perspective on childhood factors associated with poor and good health status later in life.
- The finding that Notch, a cell membrane protein involved in cell-cell communication, is also critically involved in the death of nerve cells following a stroke. These findings reveal Notch signaling as a novel therapeutic target for stroke and related neurodegenerative conditions.
- The elucidation of the metabolic effects of glucagon-like peptide-1 (GLP-1), a gut peptide that dramatically increases insulin secretion upon eating, has led to the development of the GLP-1 receptor agonist, Exendin-4, which has been approved by the FDA as a treatment for type 2 diabetic patients.
- The identification of Dardarin (LRKK2), the most common genetic cause of Parkinson's disease to date.
- Experimental animal model studies illustrating the protective antiapoptotic effect elicited by the cytokine erythropoietin on the heart when stressed by ischemia.

Training Scientists for the Future

As the NIA leads the federal effort on aging research, it is committed to training the next generation of researchers for lifetime careers in the biomedical and behavioral sciences. The Intramural Research Program offers fellowship training opportunities for both pre-doctoral and postdoctoral candidates. Opportunities exist to conduct laboratory, clinical and translational research in a vast array of biomedical disciplines. The IRP provides a stimulating academic setting for a comprehensive effort to understand aging through multidisciplinary investigator-initiated research.

<http://www.grc.nia.nih.gov/branches/osd/factbook2004.pdf>

Dan L. Longo, M.D.
Scientific Director



Interview/Hire the World's Best Scientists

Reach innovative young scientists at the forefront of their fields

Attend the Job Fair for Postdoctoral, Research, and Clinical Fellows

Date: October 19, 2006

Time: 10:30 am - 3:00 pm

Place: Natcher Conference Center
National Institutes of Health
Bethesda, MD

Access 3800 doctoral scientists and clinicians in training at the NIH

Register by September 15th at:
www.training.nih.gov/jobfair

Office of Intramural Training and Education

NIH is dedicated to building a diverse community in its training and employment programs.



NIH physician specialty training

your goal is to pursue academic medicine, and you have a U.S. medical license, consider fellowship or residency training with the physicians who are performing basic research and designing pivotal trials that will determine state-of-the-art care for the next decade. The National Institutes of Health is home to 250 research beds. We offer outstanding clinical training and a research experience that includes intense exposure to designing and analyzing clinical trials, and to cutting-edge basic science investigation, within our ACGME-accredited programs.

Student Loan Repayment of up to \$35,000 per year may be available to qualified individuals accepted into ACGME-accredited training programs.

For more information on ACGME-accredited programs and the Loan Repayment Program, visit our web sites at www.training.nih.gov, www.lrp.nih.gov

Office of Intramural Training and Education
Bethesda, Maryland 20892-1158
(888) 695-5343

NIH is dedicated to building a diverse community in its training and employment programs.

subspecialty training

Internal Medicine
Allergy and Immunology
Critical Care Medicine
Endocrinology and Metabolism
Hematology
Infectious Diseases
Medical Oncology
Rheumatology

Pediatrics
Allergy and Immunology
Endocrinology
Hematology/Oncology
Pediatrics/Medical Genetics

Blood Banking/Transfusion Medicine
Cytopathology
Hematopathology
Medical Genetics

residency training

Anatomic Pathology
Dermatology (*third year only*)
Psychiatry (*fourth year only*)

resident and medical student electives

Clinical and research electives for residents.
Flexible duration, 4 weeks, 8 weeks, up to one year.



Nadia's nectar

The fountain of youth.

Ian Watson

Nadia's Nectar poured over cubes of watermelon makes a delicious breakfast!

"We'll try that!"

Nadia Peartree was the toast of Hollywood by 2015! Two Golden Globes, and we ain't referring to her bosom, plus two Oscars — !

"Mom, I'm trying to do Tyrannosafari on my fone."

"Just eat your toast, Pumpkin, and drink your nectar."

Plus, Nadia was Born Again, so when she quoted Solomon's Proverbs 5:15, 'Drink water from your own cistern,' America listened. And bought — !

"Honey, if you'll quit fondling that T-Pak, it'll shut up."

Would you believe that at the turn of the twenty-first century only three million Chinese drank urine; and those Brahmins in India and in the West folks who were yoga fanatics — ?

"But it's a new touchy-talkchip, Bert."

Yes, folks, pee-prejudice blinded most of us to a pedigree going back to ancient Israel and Tibet, whose lamas lived to a ripe old age because of pee, which rhymes with immortality! Gandhi drank his own urine, and he beat the British —

"Honey, I'm sick of every darn food and drink voicing off — 'watch out, I expire next week, by the way tortoises can live 200 years, well I ain't no tortoise' — yabber yabber."

They say Steve McQueen lived on urine and boiled alligator skin while he was fighting the big C —

"Bert, I never knew that."

Though most likely Nadia was inspired by Paul Newman's Organic Foods, what with all his royalties after tax, hundreds of millions, going to education and charity work. Remember Pa Newman and his daughter dressed up on those packets of healthy pretzels and choc bars and alphabet cookies looking American Gothic? Wholesome traditional virtues! Funnily enough, an FAQ was, 'Do you use slave labour in producing your chocolate?' What a question! Of course not! The welfare of producers of the raw material was paramount (and I don't mean as in Pictures!). Nadia thought even more deeply about the welfare of producers in the Third World —

"Mom, I can't hear if a Tyranno's coming."

"Pumpkin, I'm just pouring myself a little more."

"Yeah, drip by drip."

This was the philanthropic clincher! Only naive buyers of Nadia's Nectar could think they were quaffing urine solely passed by the star herself — leaving aside that 'passed' also means 'approved'! Good wholesome urine has to be sourced in considerable bulk, pasteurized and frozen for shipment. Where better, than from the poverty-stricken lands of Central America? What a godsend to so many people in those nations!



"Oh zip your lip, T-pak."

"It ain't got ears or lips, Pa."

For the urine to be first rate, the donors all need to be in tip-top health, kept supplied with all necessary nutrition, vitamins and minerals in exchange for their urine. The Man from Nadia's Nectar doesn't visit joyful producers to slice open a sample fruit with a machete! No, he's there all the time with a biochem test kit to check up on everyone, and a good nose for the bouquet of quality urine. It's famous that Nadia maintains the most wonderful health clinics in the source areas.

"Just hear a bit more, Bert?"

Nadia also contributes directly to her product, even now, when she's in her late fifties, though looking a million dollars, needless to say. Into every hundred pints goes a droplet of her own urine. By homeopathic principles this intensifies the effect! All consumers know that they're participating directly in Nadia's glamour.

In the old days you'd hear silly objections that you couldn't get much in the way of urokinase or assorted hormones and antibodies and magnesium, calcium,

potassium, out of quaffing a pint of pee — and also that the body was peeing stuff out because it already had enough inside it, that it actively wanted rid of the stuff. Tell that to a two-hundred year old lama.

"How much more?"

"But Bert, we love Nadia."

"I'm drinking her, ain't I?"

Did ya know how good-ole Enzymes of America used to fit special filters to 10,000 Porta-Johns to recover an enzyme that dissolves blood clots? Fourteen million gallons of pee going into Porta-Johns each year meant quarter of a million coronary arteries unblocked! There was a half-billion-dollar annual market for urine products. But until Nadia's Nectar, that was an invisible market, a silent one. Pee products were hidden in pills and small print. Oh, and in beauty products and soaps — urine breaks down grease in hair, fr'instance.

Who cares if some other former Oscar-winners copied Nadia with rival brands, and quirky flavours and colours? For my money nothing beats Nadia's Original Nectar.

If your eyes are tired, give them a few drops of her urine. Aching ears? Likewise. Sinuses congested? Sniff the golden juice! Drink water from your own well! Or better, from Nadia's. The Bible says so, and Nadia made sure we got the message.

Oh, the early commercials were so classy! Elegantly nightgowned Nadia going into her personal bathroom in her Malibu mansion just after sunrise and closing the door while the tinkly Trout Quintet played, then coming out smiling gorgeously carrying a beaker of golden liquid, which an aide promptly rushed away to safeguard its freshness. A genuine classic. Any of you remember?

"Me, I do!"

The modern commercials are pretty neat too. Nadia does most ads personally and never uses a rejuvenated CGI clone. She's utterly for real. That's why most people prefer her urine to their own. Trust. Drinking her pee's a communion.

Is it a wrap, guys? Did I read it good?

"Goofs, they oughta edited that bit out."

"No, Bert, it's gotta be a deliberate mistake. There'll be a prize for noticing! Pumpkin, pass me your fone."

Ian Watson wrote the screen story for Spielberg's *A.I. Artificial Intelligence*. His next story collection is *The Butterflies of Memory* (PS Publishing, June 2006). His website with funny photos is at www.ianwatson.info.